DOCUMENT RESUME

ED 300 413 TM 012 409

AUTHOR	Baggaley, P. Jonathan, Ed.; And Others
TITLE	Evaluation of Educational Television.
INSTITUTION	South African Broadcasting Corp., Johannesburg.
REPORT NO	ISBN-0-86972-070-8
PUB DATE	87
NOTE	328p.
AVAILABLE FROM	South African Broadcasting Corporation, P.O. Box 8606, Johannesburg, 2000, Republic of South Africa.
PUB TYPE	Collected Works - General (020) -- Reports - Research/Technical (143) -- Books (010)
EDRS PRICE	MF01/PC14 Plus Postage.
DESCRIPTORS	Educational Assessment; *Educational Television; Evaluation Criteria; *Evaluation Methods; *Foreign Countries; Formative Evaluation; Preschool Education; Programing (Broadcast); Summative Evaluation; Teacher Education Programs; *Television Research; Theories; Young Adults
IDENTIFIERS	*South Africa

ABSTRACT

Eleven articles on the evaluation of educational television (ETV) in South Africa are provided. Under the heading "Theory" are: (1) "The Meaning of Evaluation and Its Practice" (D. Nevo); (2) "Criteria for Evaluating ETV: A Theoretical Framework" (R. Israeli); and (3) "Sources of Evaluation Criteria in Education, with Particular Reference to the Evaluation of ETV" (A. Lewy). Under "Method" are: (4) "Complexities of Summative ETV Evaluation: Possible Solutions through a Diagnostic Approach" (G. Salomon); (5) "Continual Response Analysis in TV Evaluation" (J. Baggaley); (6) "Formative Evaluation of ETV in the Early Stages of Programme Development" (A. Duby); (7) "Guidelines for Evaluating ETV-Programmes for Teacher Training" (N. J. de Jager and R. H. Young); and (8) "Evaluating Educational Television Programmes Produced for Pre-School Children" (E. Nel and P. de Vries). Under "Investigations" are: (9) "The Effect of Formal Features on the Attention Given to an ETV-Programme by South African Indian Pre-School Children" (A. Parekh); (10) "Determining the Optimal Placing for Young Adults of Activities for ETV-Programmes" (R. M. Yule); and (11) "The Evaluation of ETV-Programmes in South Africa: Some HSRC Initiatives" (D. P. van Vuuren and C. Herselman). (HSRC refers to the Human Services Research Council.) (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED300413

EVALUATION OF EDUCATIONAL TELEVISION

Editors: J Baggaley, A Duby, A Lewy

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

ALIZA DUBY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM012409

SOUTH AFRICAN BROADCASTING CORPORATION
JOHANNESBURG, 1987

BEST COPY AVAILABLE

EVALUATION OF EDUCATIONAL TELEVISION

Editors: J Baggaley, A Duby, A Lewy

THE EDITORS ACKNOWLEDGE THE ASSISTANCE OF DR P DE VRIES, DIVISION OF ADULT EDUCATION, UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG WITH THE COPY EDITING OF THIS BOOK.

First published and printed in 1987 by the South African Broadcasting Corporation, P.O. Box 8606, Johannesburg, 2000

ISBN 0 86972 070 8

Copyright : SABC, Johannesburg

The opinions expressed in the respective articles in this book do not necessarily reflect the official view of the Department of Educational Programmes RTV, or that of the SABC

Wordprocessing and layout : Magda Gildenhuys

CONTRIBUTORS

1. BAGGALEY, P. Jonathan - Professor in Department of Education, Concordia University, Montreal. Author of the book 'Psychology of the TV Image' and others books and articles on television research, Associate editor of *Journal of Educational Television*.

2. DE JAGER, Nico - Professor and Head of the Department of Teacher Training, University of Port Elizabeth. He has published articles in South African journals.

3. DE VRIES, Peter - Senior lecturer in the Division of Adult Education, University of the Witwatersrand. Editor of *ELTIC Reporter:* English Language Teaching Information Centre. His other fields of interest are English Second-Language Instruction, as well as evaluation studies.

4. DUBY, Aliza - Project leader of evaluation studies in the Educational Programmes Department at the South African Broadcasting Corporation. She has published in international journals and in Israel and in South Africa. Her fields of Research: Evaluation, Media and Games.

5. ISRAELI, Rachel - Research Fellow of Teacher Opportunity Programmes, Johannesburg. Field of specialisation Management Sciences.

6. LEWY, Arieh - Professor of Education, Tel-Aviv University. His field of Specialisation is Programme Evaluation. Editor of the *International Review of Journal Studies in Educational Evaluation.*

7. NEL, Elizabeth - Assistant Professor in the Department of Psychology, Stellenbosch University. Author of the book *US and Television*. Her field of interest is the influence of television on children.

8. NEVO, David - Head of the graduate programme of research methods and evaluation in the School of Education, Tel-Aviv University. His areas of specialisation include evaluation theory, research methodology and measurement, in which he has published numerous books and articles in professional journals.

9. PAREKH, Angina - Lecturer in the Department of Psychology, University of Durban, Westville. Her field of interest is early childhood education.

10. SALOMON, Gavriel - Professor in School of Education, Tel-Aviv University. He leads the graduate programme of communication and technology in Education. Salomon is the author of three books and numerous research and theoretical papers on the cognitive effects of media.

11. VAN VUUREN, Daan - Head of Broadcasting Research, SABC. His fields of interest are Mass Media Research and the Impact of TV on Children. He has published numerous articles in professional journals.

12. YOUNG, Richard - Senior lecturer, University of Port Elizabeth. His fields of specialisation is Didactics in History teaching and published articles on teaching history in school.

13. YULE, Ronald - Rector of Pretoria College of Education and Professor in the Department of Education, University of South Africa.

FOREWORD

J A Eksteen

DIRECTOR-GENERAL
SOUTH AFRICAN BROADCASTING CORPORATION

South Africa was a late entrant in the world of television, so we were able to benefit from the experience of other countries, and hence install a national television service of exceptional technical quality. Likewise, when the South African Broadcasting Corporation undertook to produce and broadcast non-formal educational television programmes, extensive efforts were made to learn from the experience of leading educational television centres.

To our surprise, we found that the guidelines we were seeking were not readily available in publications, and much of the published work had yet to be scientifically validated. Hence the idea was conceived of collecting together into a compact volume information relating to the criteria and guidelines for the evaluation of educational television programmes.

In keeping with its origins, efforts were made to maintain a cosmopolitan perspective in this book. So we are pleased to be able to include contributions from distinguished international researchers; to Prof Gavrial Salomon and Prof Arieh Lewy of Israel, and Prof Jon Baggaley of Concordia University, Canada, we express our sincere appreciation.

We also wish to record our appreciation to Dr P J van Zyl, who initiated this project, and to the project leader, Dr Aliza Duby, who co-ordinated the compilation of this volume.

We trust that those involved in educational television will find this book helpful and useful.

Author		Contents	Page
EKSTEEN J A DIRECTOR-GENERAL, SABC	:	FOREWORD	IV
BAGGALEY J	:	PREFACE	1

Part One: Theory

		Chapter One	
NEVO D	:	THE MEANING OF EVALUATION AND ITS PRACTICE	4
		Chapter Two	
ISRAELI R	:	CRITERIA FOR EVALUATING ETV: A THEORETICAL FRAMEWORK	24
		Chapter Three	
LEWY A	:	SOURCES OF EVALUATION CRITERIA IN EDUCATION, WITH PARTICULAR REFERENCE TO THE EVALUATION OF ETV	49

Part Two: Method

		Chapter Four	
SALOMON G	:	COMPLEXITIES OF SUMMATIVE ETV EVALUATION: POSSIBLE SOLUTIONS THROUGH A DIAGNOSTIC APPROACH	109
		Chapter Five	
BAGGALEY J	:	CONTINUAL RESPONSE ANALYSIS IN TV EVALUATION	136
		Chapter Six	
DUBY A	:	FORMATIVE EVALUATION OF ETV IN THE EARLY STAGES OF PROGRAMME DEVELOPMENT	171

Author	Contents	Page
DE JAGER N J AND YOUNG R H	: *Chapter Seven* GUIDELINES FOR EVALUATING ETV-PROGRAMMES FOR TEACHER TRAINING	222
NEL E AND DE VRIES P	: *Chapter Eight* EVALUATING EDUCATIONAL TELEVISION PROGRAMMES FOR PRE-SCHOOL CHILDREN	243

Part Three: Investigations

Author	Contents	Page
PAREKH A	: *Chapter Nine* THE EFFECT OF FORMAL FEATURES ON THE ATTENTION GIVEN TO AN ETV-PROGRAMME BY SOUTH AFRICAN INDIAN PRE-SCHOOL CHILDREN	263
YULE R M	: *Chapter Ten* DETERMINING THE OPTIMAL PLACING FOR YOUNG ADULTS OF ACTIVITIES FOR ETV-PROGRAMMES	284
VAN VUUREN D P AND HERSELMAN C	: *Chapter Eleven* THE EVALUATION OF ETV-PROGRAMMES IN SOUTH AFRICA: SOME HSRC INITIATIVES	303
LEWY A	: CONCLUDING COMMENTS	317

PREFACE

J Baggaley

In more ways than one, a TV producer is like an architect. He creates a complicated structure in which others may find relaxation or enlightenment. He must ensure that the basis for the structure is solid. He must also be responsive to those who will use it. The architect has developed this process to a fine art. He presents the plans to his client for approval, and defends them when changes become necessary. He submits the structure to repeated tests to ensure that it does not fail when used. He knows that if the building fails to meet his client's objectives, he will be held accountable[1].

The failure of an educational TV programme, however, can go completely unnoticed. If it is entertaining, a film is apt to be judged successful whether or not its viewers are edified, and the quality of their lives improved. Even in broadcasting organizations, with facilities for audience research, checks on the educational impact of programming are the exception rather than the rule.

In the past, this has largely been due to the shortcomings of audience research methodology. Traditional audience research methods have yielded general assessments of audience impact with implications for programme scheduling, for the evaluation of production personnel, and for overall broadcast policy. Lacking immediacy and precision, however, they have generated few insights for producers to apply in practice – in seeking to improve a programme, for example, or to make a new one.

---

1) I am personally indebted, for the 'film-maker as architect' analogy, to Kirk Smith, a Canadian film-maker, with whom I have conducted several projects.

The traditional methods of audience research are far more likely, therefore, to be used against a producer's interests, rather than for them; and it is hardly surprising that today's media producers are inclined to distrust evaluation and research efforts, and to resist them actively.

Since the 1970's, however, new methods and perspectives in media research have led to a greater understanding of media communication techniques, and of audience reactions to them. The educational media industry has been slower to benefit from these developments than the advertising industry. The aim of this book is to help redress this balance. Criteria have been assembled for the education producer to consider when designing material for children, for adult education purposes, and in teacher training. A range of steps are shown - some simple, others more complicated - that can be taken to monitor and ensure the success of an educational production; and the problems of successful collaboration between media producers and researchers in the pressured production environment are discussed. Above all, the book aims to help TV and film producers to gain the degree of control which an architect applies in the construction of a house.

The book draws heavily on the example set by the Children's Television Workshop of New York. Many of the specific findings and approaches discussed in the book derive from the in-depth collaboration of researchers and producers on CTW productions, such as *'Sesame Street'* and *'3-2-1 Contact'*. However, the book is essentially the product of an unusually fertile environment at another educational production house, that of the South African Broadcasting Corporation (SABC) in Johannesburg. Television is still a newcomer to South Africa (introduced in 1976); and educational television services began at the SABC only in 1984. In that time, a healthy spirit of co-operation has been apparent between SABC's producers and researchers - possibly because the organization has been able to anticipate many of the problems previously reported by producers and researchers elsewhere.

In the last two years, the researchers of South African broadcasting have reported an impressive range of studies, relating to TV productions for the country's diverse social population groups. One such study guided the production of *'Siyafunda'*, a series which has taught the Zulu language to all of South Africa's population groups, so that it may be available as a common language for the whole country (Landers, 1986). Another study in progress is using the most advanced programme evaluation techniques in the development of the first Zulu 'soap opera'. It has been my pleasure to be acquainted with these valuable cross-cultural projects at first hand.

As co-editor of this book, I have depended heavily on my colleague Dr Aliza Duby, who has co-ordinated the project throughout its two-year inception. We have both been grateful for the scholarship and guidance of our fellow editor, Professor Arieh Lewy. All three editors thank the contributing authors, who have surmounted the barriers of time and space with willingness and enthusiasm.

REFERENCES

Baggaley, J P and Smith, K 1982. Formative research in rural education. Media in Education and Development 15, pp 173-6.

Landers, M 1986. Formative evaluation of 'Siyafunda'. 6th International Conference on TV Research, Montreal.

Part One: Theory

THE MEANING OF EVALUATION AND ITS PRACTICE

D Nevo

Many attempts have been made in recent years to clarify the meaning of evaluation and expose the distinction between evaluation and other related concepts such as measurement and research. The literature contains many approaches to the conceptualization of evaluation, and the determination of its countenance in education. Many of those approaches have been incorrectly referred to as "models" - e.g. the CIPP Model, the Discrepancy Model, the Responsive Model or the Goal-Free Model - in spite of the fact that none of them includes a sufficient degree of complexity and completeness that might be suggested by the term "model". Stake (1981) has suggested that they be referred to as persuasions rather than models.

For the benefit of those of us who have lost our way among the various evaluation models, approaches and persuasions, several attempts have been made to put some order into the growing evaluation literature by classifying evaluation approaches. Such classifications (e.g. Worthen and Sanders, 1973; Popham, 1975; Stake, 1976; Stufflebeam and Webster, 1980; House, 1980; Guba and Lincoln, 1981; Madaus et al, 1983) have made a significant contribution through their critical reviews of evaluation literature, noting similarities and differences among the various approaches. However, those classifications have treated each model as a whole; moreover the categorization schemes have not always brought out the major issues underlying the agreements and disagreements among the various approaches.

Stufflebeam (1974) suggested eight questions to be addressed in any attempt to conceptualise evaluation.

Nevo (1981, 1983) revised Stufflebeam's list of questions and extended it to a list of ten major aspects in a conceptualization of evaluation. These ten represent the major issues addressed by the most prominent evaluation approaches in education. They will be used here as an organizing tool for an analytical review of the literature on educational evaluation, and as a framework for our attempt to understand the concept of evaluation.

THE TEN ASPECTS ARE:

- The definition of evaluation
- Functions of evaluation
- Evaluation objects
- Evaluation variables
- Evaluation criteria
- Evaluation clients and audiences
- The evaluation process
- Evaluation methods of inquiry
- Evaluators
- Evaluation standards

The Definition of Evaluation

Many definitions of evaluation can be found in the literature. In his well-known definition, Ralph Tyler perceives evaluation as "the process of determining to what extent the educational objectives are actually being realized" (Tyler, 1950:69). Another widely accepted definition of evaluation has been that of providing information for decision making, suggested by various leading evaluators such as Cronbach (1963), Stufflebeam (Stufflebeam et al, 1971) and Alkin (1969). In recent years consensus has been reached among evaluators on the definition of evaluation as the assessment of merit or worth (Scriven, 1967; Glass, 1969; Stufflebeam, 1974; Eisner, 1979; House, 1980), or as an activity comprising both description and judgement (Stake, 1967; Guba and Lincoln, 1981). A joint committee on standards for evaluation (see below) recently published its definition as "the systematic investigation of the worth or merit of some object" (Joint Committee, 1981:12).

A major departure from the consensus, the judgemental definition of evaluation is the Standford Evaluation Consortium group's definition of evaluation as "(a) systematic examination of events occurring in and consequent on a contemporary program - an examination conducted to assist in improving this program and other programs having the same general purpose" (Cronbach et al, 1980:14). Cronbach and his associates clearly reject the judgemental nature of evaluation by advocating an approach which perceives the evaluator as "an educator (whose) success is to be judged by what others learn" (Cronbach et al, 1980:11) rather than a "referee (for) a basketball game" (p 18) who is hired to decide who is "right" or "wrong".

A definition which points to the judgemental character of evaluation might create a considerable amount of anxiety among potential evaluees and raise resistance among opponents of evaluation. Obviously, a non-judgemental definition of evaluation - such as "providing information for decision making" - might be accepted in a more favourable light by evaluees and clients. However, it may be unrealistic to create positive attitudes towards evaluation by minimizing one of its major features. Another approach intended to develop positive attitudes towards evaluation might be to demonstrate its constructive functions (e.g. formative evaluation) within the various domains of education, as can be seen in the next section.

The Functions of Evaluation

Scriven (1967) was the first to suggest the distinction between "Formative Evaluation" and "Summative Evaluation", referring to two major functions of evaluation, although he was not the first to realise the importance of such a distinction. Later on, referring to the same two functions, Stufflebeam (1972) suggested the distinction between pro-active evaluation intended to serve decision-making, and retro-active evaluation to serve accountability. In its formative (pro-active) function, evaluation is used for the improvement and development of an on-going activity (or programme, person, product, etc.). In its summative (retro-active) function, evaluation is used for accountability, certification, or selection.

A third function of evaluation, which has been less often treated by evaluation literature (House, 1974; Patton, 1978; Cronbach et al, 1980) is the psychological or socio-political function. In many cases it is apparent that evaluation is not serving any formative purposes, nor is it being used for accountability or other summative purposes. Instead, it is being used to increase awareness of special activities, motivate desired behaviour of evaluees or promote public relations. Regardless of personal feelings about the use (or misuse) of evaluation for this purpose, this function cannot be ignored.

Another somewhat "unpopular" function of evaluation is its use for the exercise of authority (Dornbush and Scott, 1975). In formal organizations, it is the privilege of the superior to evaluate his or her subordinates, and not vice versa. In many cases a person in a management position might use evaluation processes as a demonstration of his authority over a sub-ordinate. This may be referred to as the "administrative" function of evaluation.

To summarise, evaluation can serve many functions: (a) the formative function - for improvement, (b) the summative function - for selection, for certification, for accountability, (c) the psychological or socio-political function - for motivation, to increase awareness, and (d) the administrative function - to exercise authority.

Although some evaluators (Alkin et al, 1979; Cronbach et al, 1980) express a clear preference for the formative function of evaluation, the general view seems to be that there are no "right" or "wrong" roles, and that more than one function can be consciously served by evaluation. However, each different evaluation function may require different methods of implementation. It is therefore, very important to be aware of the differences, and to determine the specific functions of an evaluation function at a very early stage of planning.

Evaluation Objects

Students and teachers have always been popular 'objects' of evaluation in education. Almost all of the measurement and evaluation literature in education up to the mid-sixties dealt with the evaluation of students' learning. Up to that time one could hardly find substantial guidance on the evaluation of other objects such as educational projects, programmes, curricular materials or educational institutions. Developments in the educational system of the United States (e.g. the Elementary and Secondary Education Act of 1965) led to a significant shift of focus in the objects of educational evaluation from students to projects, programmes and instructional materials, and the latter objects became the common topic of the evaluation literature of education (Stake, 1967; Scriven, 1867; Stufflebeam, 1969; Alkin, 1969; Provus, 1971; Stufflebeam et al, 1971).

In planning an evaluation, it is very important to determine the limits of "the thing" (or "the evaluand" to use Scriven's (1980) term) that has to be evaluated. A clear identification of the object helps keep an evaluation focussed. It also helps to clarify who will be affected by the evaluation, to resolve value conflicts among stakeholders and to avoid the potential threat among others involved in the evaluation (Guba and Lincoln, 1981). It helps to determine what kind of information should be collected and how it should be analysed.

Evaluation Variables

After an evaluation object has been chosen, a decision has to be made as to which aspects and characteristics of that object will be evaluated. Information pertinent to such aspects has to be collected. Earlier approaches to evaluation focussed mainly on results or outcomes. In that case, to evaluate an educational object (e.g. a new curriculum) meant evaluating the quality of the results of its functioning (e.g. students' achievements).

In recent years some interesting attempts have been made to extend the scope of evaluation variables in various evaluation models (Stake, 1967; Stufflebeam, 1969, 1974; Stufflebeam et al, 1971; Alkin, 1969; Provus, 1971). Stufflebeam's CIPP Model suggests that evaluation focus on four variables for each evaluation object: (1) its goals, (2) its design, (3) its process of implementation, and (4) its outcomes. According to this approach, the evaluation of an educational project, for example, would assess (a) the merit of its goals, (b) the quality of its plans, (c) the extent to which those plans are being carried out, and (d) the worth of its outcomes.

Stake (1967) suggested in his "Countenance Model" that two sets of information be collected on the evaluated object: descriptive and judgemental. The descriptive set should focus on intents and observations of antecedents (prior conditions that may effect outcomes), transactions (the process of implementation), and outcomes. The judgemental set of information comprises standards and judgements of the same antecedents, transactions and outcomes.

Guba and Lincoln (1981), expanding Stake's Responsive Evaluation Model (Stake, 1975) and applying the naturalistic paradigm, suggest that the evaluator generate five kinds of information: (1) descriptive information of the evaluation object, its setting, and its surrounding conditions, (2) information responsive to concerns of relevant audiences, (3) information about relevant issues, (4) information about values, and (5) information about standards relevant to worth and merit assessments.

Thus, the evaluation literature seems to suggest that a wide range of information on the evaluated object can be collected through the evaluation process. It should not limit itself to the narrow scope of evaluation of outcomes or results. This does not mean that in each evaluation all possible kinds of information must always be collected; the focus should be, rather, on the kinds of information which are relevant to the identified evaluation priorities, or practical constraints.

Evaluation Criteria

Choosing the criteria for judging the merit of an evaluation object is one of the most difficult tasks in educational evaluation. Those who think that evaluation should attempt to determine whether goals have been achieved (Tyler, 1950; Provus, 1971) make this task easy for themselves by partially ignoring the issue of evaluation criteria. What they actually do is use "goal achievement" as the evaluation criterion, without having justified the goal as an appropriate criterion. What about trivial goals or all kinds of "stated objectives" that are not worth achieving? Should they be used as evaluation criteria?

Another way to "avoid" the issue of evaluation criteria is to ignore the judgemental nature of evaluation. Those who defined evaluation as an information collection activity to serve decision-making or other purposes (Cronbach, 1963; Stufflebeam, 1969; Alkin, 1969), do not have to deal with the problem of choosing evaluation criteria.

The achievement of (important!) goals is one possible basis for evaluation criteria. Alternative bases for evaluation criteria suggested by the literature might be: identified needs of actual and potential clients (Scriven, 1972; Patton, 1978; Joint Committee, 1981), ideals or social values (House, 1980; Guba and Lincoln, 1981), known standards set by experts (Stake, 1967; Eisener, 1979; Guba and Lincoln, 1981) or the quality of alternative objects (Scriven, 1967; House, 1980).

There seems to be agreement among most evaluation experts that the criteria to be used for the assessment of a specific object must be determined within the specific context of the object and the function of its evaluation. While in many cases the evaluator does not have the authority to choose among the various alternative criteria, it is the evaluator's responsibility to ensure that such a choice be made, and that he provide a sound justification for the choice made by him or by somebody else.

Evaluation Clients and Audiences

If evaluation is to be useful at all, it has to be useful to some specific client or audience. The evaluation literature does not suggest which is the "most appropriate" audience for evaluation, but each position entails its own "appropriate" clients.

Those who define evaluation as providing information for decision-making (Cronbach, 1963; Alkin, 1969; Stufflebeam et al, 1971), identify the relevant decision-makers and attempt to determine their information needs. Others (House, 1980; Cronbach et al, 1980) reject the notion of serving "decision-makers" because of the threat of co-optation or over-simplification of social and organizational processes. Cronbach and his associates (Cronbach et al, 1980) are inclined to serve the "Policy Shaping Community" rather than some kind of managerial decision-maker. Many authors refer to "evaluation clients" or "evaluation audiences" as those who have to be served by evaluation. Guba and Lincoln (1981) suggested the term "stakeholders" or "stakeholding audience" for the whole group of persons having some stake in the performance of the evaluand and who, therefore, should be served by the evaluation.

Three important propositions can be found in writings regarding this issue. They are that

- an evaluation can have more than one client or audience;
- different evaluation audiences might have different evaluation needs; and
- the specific audiences for an evaluation and their evaluation needs have to be clearly identified at the early stages of planning an evaluation.

Differences in evaluation needs might be reflected in many ways: e.g. the kind of information to be collected, the level of data analysis to be used or the form of reporting the evaluation results. Sometimes it is impossible to serve all the identified evaluation needs at the same time, and some priorities have to be set for the specific evaluation needs of the evaluation.

The Evaluation Process

The process of an evaluation might differ according to the theoretical perception guiding the evaluation. A theoretical approach in which evaluation is perceived as an activity for determining whether goals have been achieved (Tyler, 1950), might have the following evaluation process: (a) stating goals in behavioural terms, (b) developing measurement instruments, (c) collecting data, (d) interpreting findings, and (e) making recommendations.

According to Stake's "Countenance Model" (Stake, 1967), the evaluation process should include: (a) describing a programme, (b) reporting the description to relevant audiences, (c) obtaining and analysing their judgements, and (d) reporting the analysed judgements back to the audiences. Later on, in his Responsive Evaluation Model, Stake (1975) suggested a continuing "conversation" between the evaluator and all other parties associated with the evaluand. He specified twelve steps of dynamic interaction in the process of conducting an evaluation between the evaluator and his audiences.

Provus (1971) proposed a five-step evaluation process including: (a) clarification of the programme design, (b) assessing the implementation of the programme, (c) assessing its interim results, (d) assessing its long-term results, and (e) assessing its costs and benefits.

The Phi Delta Kappa Study Committee on evaluation (Stufflebeam et al, 1971) presented a three-step evaluation process. It included: (a) delineating information requirements through inter-action with the decision-making audiences, (b) obtaining the needed information through formal data collection and analysis procedure, and (c) providing the information to decision-makers in a communicable format.

Scriven 1972(a)) has suggested nine steps in his "Pathway Comparison Model". Guba and Lincoln (1981) suggest that a naturalistic-responsive evaluation be implemented through a process including the following four stages: (a) initiating and organizing the evaluation, (b) identifying key issues and concerns, (c) gathering useful information, and (d) reporting results and making recommendations.

While there seems to be no agreement among evaluation experts on the "best" process to follow when conducting an evaluation, most of them would agree that all evaluations should include a certain amount of inter-action between evaluators and their audiences at the outset of the evaluation to identify evaluation needs, and at the conclusion to communicate findings. Evaluation cannot be limited to the technical activities of data collection and analysis.

Evaluation Methods of Inquiry

While challenging the usefulness of various research methods for evaluation studies, Guba, 1969 and Stufflebeam et al, 1971, have recently also introduced a variety of methods of inquiry into the field of educational evaluation. In addition to traditional experimental and quasi-experimental designs (Campbell, 1969; Stanley, 1972; Cook and Campbell, 1976), naturalistic methods (Patton, 1980; Guba and Lincoln, 1981), jury trials (Wolf, 1975), case studies (Stake, 1978), art criticism (Eisner, 1979), journalistic methods (Guba, 1978), the modus operandi method (Scriven, 1974), and many others have become legitimate strategies in the process of evaluation.

Methodologists still propogate the superiority of certain methods: from experimental design (Rossi et al, 1979; Boruch and Cordray, 1980), at one extreme end of the spectrum, to naturalistic methods (Guba and Lincoln, 1981; Patton, 1980; House, 1980), at the other. Overall, however, there seems to be support for a more eclectic approach to evaluation methodology. At the present state of the art "the evaluator will be wise not to declare allegiance to either a quantitative - scientific - summative methodology or a qualitative - naturalistic - descriptive methodology" (Cronbach et al, 1980, p 7). However, the optimal procedure for answering a particular evaluation question should be to use the best set of methods rather than assuming that one method is best for all purposes.

Evaluators

On becoming a professional group, evaluators devoted a lot of attention to identifying the characteristics of "good" evaluators and appropriate ways to train them (Stufflebeam et al, 1971; Worthen, 1975; Cronbach et al, 1980; Boruch and Cordray, 1980; Guba and Lincoln, 1981). To be a competent and trustworthy evaluator, one needs to possess a wide variety of characteristics. These characteristics include: technical competence in the area of measurement and research methods, understanding of the social context and the substance of the evaluation object, human relations skills, personal integrity, objectivity, as well as characteristics related to organizational authority and responsibility. As it is difficult to find one person with all those qualifications, it often becomes necessary to create an evaluation team, or choose the person with the most appropriate characteristics for a specific evaluation task.

Two important distinctions have been suggested for choosing an evaluator: the first is the distinction between an internal evaluator and an external evaluator (Scriven, 1967, 1975; Stufflebeam et al, 1971; Stake and Gjerde, 1974). An internal evaluator of a project is usually one who is employed by the project and reports directly to its management.

His objectivity, as well as his external credibility, is likely to be lower than that of an external evaluator who is not directly employed by the project and enjoys a higher degree of independence.

The second distinction is between a professional evaluator and an amateur evaluator. This distinction, suggested by Scriven (1967), refers to two different foci of training and expertise rather than to a value judgement of the quality of an evaluation. An amateur evaluator is usually one whose major professional training is not in evaluation, and his involvement in evaluation represents only part of his job description (e.g. the associate director of a new maths curriculum development project conducting the formative evaluation of the project; he has an M A in maths education and some on-the-job-training in evaluation).

A professional evaluator is one with extensive training in evaluation and whose major (or even only) responsibility is conducting the evaluation (e.g. the internal evaluator of a special educational project; she has an M A in Measurement and Evaluation and five years experience in evaluating special educational projects). While the amateur evaluator's technical evaluation skills might be lower than those of a professional evaluator, he might have a better understanding of the unique evaluation needs of the project and be able to develop better rapport with members of the project. In addition, there may be an internal-amateur evaluator, and external-amateur, and internal-professional evaluator, etc.

Evaluation Standards

Several attempts have been made in recent years to develop standards for evaluating educational and social programmes (Stufflebeam et al, 1971; Tallmadge, 1977; U.S. General Accounting Office, 1978; Rossi, 1982; Joint Committee, 1981). While some writers (Cronbach et al, 1980; Stake, 1981) have criticised standard-setting in evaluation as being premature at the present state of the art, there seems to be a great deal of agreement on the need to define the scope and content.

Boruch and Cordray (1980) analysed six sets of such standards: they found that the standards were similar. The most elaborate and comprehensive set of standards, based on the largest amount of consensus, is probably the set developed and published by The Joint Committee on Standards for Educational Evaluation (Joint Committee, 1981). These standards have been developed by a committee of seventeen members, chaired by Dr Daniel Stufflebeam. The committee represented twelve professional organizations associated with educational evaluation. Thirty standards divided into four major groups have been suggested by the committee: utility standards (to ensure that evaluation serves practical information needs), feasibility standards (to ensure that evaluation is realistic and prudent), propriety standards (to ensure that evaluation is conducted legally and ethically), and accuracy standards (to ensure that evaluation reveals and conveys technically adequate information).

Summary

At the risk of over-simplification, one could summarise the review of the literature with the following ten points (describing the current understanding of the meaning of educational evaluation). The ten points are as follows:

* Educational evaluation is a systematic description of educational objects and/or an assessment of their merit or worth.

* It can serve four different functions: formative (for improvement), summative (for selection and accountability), socio-political (to motivate and gain public support), and administrative (to exercise authority).

* Any entity can be an evaluation object. Typical evaluation objects in education are: students, educational and administrative personnel, curricula, instructional materials, programmes, pojects and institutions.

- Four groups of variables should be considered for each object. They focus on the goals of the object, its strategies and plans, its process of implementation, and its outcomes and impacts.

- The following criteria should be considered in judging the merit or worth of an educational object: response to identified needs of actual and potential clients; achievement of national goals, ideals or social values; meeting agreed-upon standards and norms; comparison with alternative objects, and achieving (important!) stated goals of the object. Multiple criteria should be used for evaluating any object.

- Evaluation should serve the information needs of all actual and potential parties interested in the evaluation object ("stakeholders"). It is the responsibility of the evaluator(s) to delineate who the actual stakeholders of an evaluation are, and identify or project their information needs.

- Regardless of its method of inquiry, an evaluation process should include the following three activities: focusing the evaluation problem, collecting and analysing empirical data, and communicating findings to evaluation audiences. There is more than one appropriate sequence for implementing those activities, and any such sequence can (and somtimes should) be repeated several times during the lifespan of an evaluation study.

- As a complex task, evaluation needs to mobilise many alternative methods of inquiry from the behavioural sciences and related fields of study, and utilise them according to the nature of a specific evaluation problem. At the present state of art an a priori preference for any specific method of inquiry is not warranted.

- Individuals or teams conducting the evaluation should have the following characteristics: extensive competence in research methodology and other data analysis techniques, an understanding of the social context and the unique substance of the evaluation object, an ability to maintain correct human relations and develop rapport with individuals and groups involved in the evaluation, and a conceptual framework to integrate all the above-mentioned capabilities.

- Evaluation should strive for an optimal balance in meeting standards of: utility (to be useful and practical); accuracy (to be technically adequate); feasibility (to be realistic and prudent); and propriety (to be conducted legally and ethically).

Conclusion

As stated at the beginning of this article, a critical analysis of the various theoretical approaches to educational evaluation might have important implications for the practitioner of evaluation, as well as for the theoretician and researchers who are concerned with developing new concepts and better methods. Both the review of the literature and the ten issues presented as an analytical framework can benefit either party.

For example, evaluators can use the framework to organise their own perceptions of the various aspects of evaluation, rather than adopt uncritically one model or another. Secondly, evaluators may encounter considerable difficulties when their perceptions of a concrete evaluation differ from those of their clients and audiences. Being aware of those differences and being able to articulate them before any evaluation project has been decided upon, may assist the evaluator avoid later confusion.

An important step is therefore to find out what various parties involved in the evaluation mean by the term evaluation; what purpose is it intended to serve, what is to be evaluated, what are the feasible alternatives for doing it, and by what standards might the evaluation be judged if it is to be conducted at all. In other words, addressing the ten issues discussed in this article might help evaluators develop a clear perception of evaluation and get a better grasp of their evaluation problems before they get themselves into all kinds of dubious evaluation adventures.

Discussions among evaluation theoreticians make a fruitful contribution to the advancement of evaluation theory and practice. Their contributions could be even more fruitful if the discussions focussed on issues in disagreement rather than on competing models and paradigms. The contribution would be even more robust if the various theoretical propositions were substantiated by research findings. The dimensions of evaluation reviewed in this article could provide a framework delineating research variables for an empirical study of evaluation. Data on how the research variables relate to one another, as well as their relationships to other variables (e.g. evaluation utilisation or variables reflecting the context of evaluation), would be very much appreciated by evaluation professionals.

REFERENCES

Alkin, M C 1969. Evaluation theory development. Evaluation Comment, 2, 2-7.

Alkin, M C, Daillak, R and White, P 1979. Using Evaluations: Does Evaluation Make a Difference? Sage, Beverly-Hills, California.

Boruch, F R and Cordray, D S 1980. An Appraisal of Educational Program Evaluations: Federal, State and Local Agencies. Northwestern University, Evanston, Ill.

Campbell, D T 1969. Reforms as experiments. American Psychologist, 24, 409-429.

Campbell, D T and Stanley, J C 1963. Experimental and quasi-experimental designs for research on teaching. N L Gage (ed.), Handbook of Research on Teaching. Rand McNally, Chicago.

Cook, T D and Campbell, D T 1976. The design and conduct of quasi-experiments and true experiments in field settings. M D Dunnette (ed.), Handbook of Industrial and Organizational Psychology. Rand McNally, Chicago.

Cronbach, L J 1963. Course improvement through evaluation. Teachers College Record, 64, May 1963:672-683.

Cronbach, L J, Ambron, S R, Dornbusch, S M, Hess, R D, Hornik, R C, Phillips, D C, Walker, D E and Weiner, S S 1980. Toward Reform of Programe Evaluation. Jossey-Bass, San Francisco.

Dornbusch, S M and Scott, W R 1975. Evaluation and the Exercise of Authority. Jossey-Bass, San Francisco.

Eisener, E W 1977. On the uses of educational connoisseurship and educational criticism for evaluating classroom life. Teachers College Record, 78, February 1977.

Eisener, E W 1979. The Educational Imagination. Macmillan, New York.

Glass, G V 1969. The Growth of Evaluation Methodology. Research paper No. 27, Boulder, Co., Laboratory of Educational Research, University of Colorado, (mimeo).

Guba, E G 1969. The failure of educational evaluation. Educational Technology, 9, 29-38.

Guba, E G 1978. *Metaphor Adaptation Report: Investigative Journalism*. Research on Evaluation Project Monograph. Northwest Regional Educational Laboratory, Portland, Oregon.

Guba, E G and Lincoln, Y S 1981. *Effective Evaluation*. Jossey-Bass, San Francisco.

House, E R 1974. *The Politics of Educational Innovation*. McCutchan, Berkley.

House, E R 1980. *Evaluating with Validity*. Sage, Beverly-Hills, California. Joint Committee on Standards for Educational Evaluation 1981. Standards for Evaluations of Educational Programs, Projects and Materials. McGraw-Hill, New York.

Lewy, A (ed.) 1977. *Handbook of Curriculum Evaluation*. Longman, New York.

Lewy, A and Nevo, D (eds.) 1981. *Evaluation Roles in Education*. Gordon and Breach, London.

Madaus, G F, Scriven, M and Stufflebeam, D L (eds.) 1983. *Evaluation Models*. Kluwer-Nijhoft, Boston.

Nevo, D 1981. The evaluation of a multi-dimensional project. A Lewy et al, *Decision Oriented Evaluation in Education: The Case of Israel*. International Science Services, Philadelphia.

Nevo, D 1983. The conceptualization of educational evaluation: an analytical review of the literature. *Review of Educational Research*, Vol. 53(1).

Patton, M Q 1978. *Utilization Focused Evaluation*. Sage, Beverly-Hills, California.

Patton, M Q 1980. *Qualitative Evaluation Methods*. Sage, Beverly-Hills, California.

Popham, W J 1975. Educational Evaluation. Prentice-Hall, Englewood Cliffs, N.Y.

Provus, M M 1971. Discrepancy Evaluation. McCutchan, Berkley, California.

Rossi, P H, Freeman, H E and Wright S R 1979. Evaluation: A Systematic Approach. Sage, Beverly-Hills, California.

Rossi, P H (ed.) 1982. Standards for Evaluation Practice. Jossey-Bass, San Francisco.

Scriven, M 1967. The methodology of evaluation, R E Stake (ed.), Curriculum Evaluation, AERA Monograph Series on Evaluation, No. 1. Rand McNally, Chicago.

Scriven, M 1972. Pros and cons about goal-free evaluation. Evaluation Comment, 3, (4).

Scriven, M 1972(a). The Pathway Comparison Model of Evaluation. Jan. (mimeo).

Scriven, M 1974. Maximizing the power of casual investigations: the modus operandi method. W J Popham (ed.), Evaluation in Education. McCutchan, Berkley, California.

Scriven, M 1975. Evaluation Bias and Its Control. Occasional Paper No. 4. Kalamazoo, Michigan: The Evaluation Center, Western Michigan University.

Scriven, M 1980. Evaluation Thesaurus (2nd ed.). Edge-press, Inverness, California.

Stake, R E 1967. The countenance of educational evaluation. Teachers College Record, 68, 523-540.

Stake, R E (ed.) 1975. Evaluating the Arts in Education: A Responsive Approach. Merrill, Columbus, Ohio.

Stake, R E 1976. Evaluating Educational Programmes: The Need and the Response. OECD Publications Center, Washington, D.C.

Stake, R E 1978. The case study method in social inquiry. Educational Researcher, 7, 5-8.

Stake, R E 1981. Setting standards for educational evaluators. Evaluation News, 5(2).

CRITERIA FOR EVALUATING ETV: A THEORETICAL FRAMEWORK

R Israeli

Systematic evaluation of educational television programmes began in the mid-1960's. Up to that time, studies in educational television had been concerned mainly with a comparison between live and televised instruction. As this line of research did not yield significant results (Chu and Schramm, 1967), it did not enhance the status of evaluation of television productions. More recently, the debate about the advantages of educational television, as well as the call for accountability have precipitated a greater demand for evaluation. As a result, most major projects using educational television today have evaluation procedures (Williams and Monty, 1976).

A greater awareness of the value of evaluating TV programmes was heightened by the evaluation of the "Sesame Street" project (Lesser, 1974) which showed that this project had been a success.

ETV-programme production calls for a joint effort by experts from widely disparate fields: subject matter specialists, educationalists, media experts, technical staff, artists, producers and directors are all expected to co-operate in producing a programme. It follows that a good ETV-programme contains features from various disciplines. Consequently, evaluation criteria should reflect the norms in these disciplines for establishing worth.

The methodological traditions of educational evaluation, instructional media studies and film criticism differ from one other; therefore their approaches to evaluating ETV-programmes will also differ. This paper suggests a way to integrate the methodological traditions of these disciplines. In the following sections a description will be given of the major features of the three approaches, followed by a suggested method for integrating them.

The Educational Evaluation Approach

Evaluation of educational programmes as a discipline emerged in the 1950's. Tyler's 'achievement of desired outcomes' (1950) was the first well-known model; subsequently, the literature on different approaches to evaluation has grown substantially. The CIPP model (Stuffelbeam, 1972), the Discrepancy model (Provus, 1971), the Countenance and Responsive models (Stake, 1967, 1981), the Goal-Free Evaluation model (Scriven, 1972), and the Illuminative Approach (Parlett, 1974) are only a few of the major models developed in this field during the last 35 years.

In his comparison of different models of educational evaluation, Nevo (1983) focuses on the ten most prominent issues: he considers the choice of "criteria for judging the merit of an evaluation object as one of the most difficult tasks in educational evaluation" (p 121).

In traditional schooling, the most important criterion for evaluation has been the acquisition of pre-specified knowledge and pre-specified moral behaviours. Contemporary education has not moved too far from this traditional approach to evaluation. Scholars in education have found that focusing only on outcome criteria unduly narrows the universe of those programme characteristics which should be considered for judging the quality of the programme.

In the last three decades, a variety of additional criteria have been identified from evaluation literature: the most important is probably the evaluation of the goals of the programme. Most evaluators have gleaned their criteria from particular theories related to the school and school curricula. For example, Tyler (1950) derived criteria for evaluation from definitions of the learners' needs, society's needs and the subject matter being taught. By contrast, Bloom (1971) derived criteria from the taxonomies suggested by him and his associates.

The integration of programme evaluation in education with evaluation activities in other fields of social action (public health, alcoholism, crime prevention, etc.) has resulted in a broadening of the classification criteria. These newly defined classifications lack unique reference to educational theories and concepts; consequently, they are characterized by a high level of generality. They may be applied equally well to evaluating educational television programmes as to evaluating other social actions such as an urban renewal project, etc. For example, Stuffelbeam (1972) specified generalized criteria classifications. He defined four major classifications and described them in terms of key words: context, input, process, product (CIPP). Another content-free classification scheme was put forward by Eraut (1975) who spoke about intrinsic versus non-intrinsic criteria. He denoted intrinsic criteria as those which could be analysed solely from the programme plan or the material used in the programme before its actual implementation. (For example; a fair representation of certain values, technical qualities, compliance with well-established learning theories, etc.)

Lewy (1977) suggested three classification criteria: matching established standards, eliciting the desired process, yielding the desired outcomes. Although these categories were derived for school curricula, they are stated in terms which are also applicable to social actions outside the education domain.

Summarizing the literature dealing with educational evaluation models, Nevo (1983) suggests five broad kinds of criteria:

- *responding to identified needs of actual and potential clients;*
- *achieving national goals, ideals or social values;*
- *meeting agreed standards and norms;*
- *out-doing alternative objects;*
- *achieving important stated goals of the object (p 125).*

To apply educational evaluation criteria usefully to ETV and assign appropriate importance to each criterion, Aversa and Morrison (1978) proposed three design variables which define the nature of the programme: instructional intent, the audience, the intended use. The instructional intent of a specific programme can be found in answers to questions such as: "Is the programme designed as a part of a formal education course or is it meant to function outside the realm of formal education?" Does it aim to inform and entertain the viewer without measuring knowledge gains or is the learner to be tested on his achievements, etc. The audience issue is important because different programmes are intended to appeal to different audiences.

A programme can be developed for a small selected audience (such as students on a study course, or people interested in a particular skill such as photography, etc.) or for a large and diverse audience. The audience itself can be either voluntary or captive and these differences should influence the choice of criteria by which the programme is evaluated.

The third design variable is its intended use. The use of this variable suggests that programmes developed for broadcasting should be judged differently from those not intended for broadcasting.

The Instructional Media Approach

Several education and media researchers have tried to differentiate between various kinds of media by offering various classifications or media-taxonomies. The most general type of classification is based on the technical qualities of the media. Bretz (1971) and Duncan (1969), using the technical-qualities approach to compare television with other media, found that television (and film) were in the highest category of media. Therefore, it can be concluded that television which has the characteristics of audio-motion-visual media has a high degree of complexity.

As the technical qualities of classifications do not throw light on the value of television as an instructional device, they have little bearing on the field of instruction. More sophisticated kinds of classification which were derived from the educational domain were suggested by Kemp (1971), Boucher et al (1973), and others. They analysed the requirements of the learning tasks and the special attributes of each medium and suggested a match between the media and tasks to be accomplished. As the attributes they considered comprised mainly the technological development and transmission facilities of the media, television may be considered as an appropriate medium for almost all learning tasks (Leech, 1980). Heidt (1980) states that "All attempts, however, to extract generalizable conclusions from these studies have failed so far .... Nevertheless, researchers as well as media practitioners agree that there are differences between different types of media, that these, however, are only relative and of a more subtle kind than hitherto expected" (p 45). He suggests that in order to identify the way in which these subtle differences affect learning gains, the gains would have to be linked to the unique attributes of each medium and not to the medium only.

This calls for a classification of media attributes rather than a classification of the media themselves. Totsi and Ball (1969) offered such a classification in which they differentiated between the medium and its presentation form. Claiming that only the form of presentation had an effect on learning, they identified several dimensions of presentation variables, such as stimulus duration and stimulus encoding form. According to this approach, different learning tasks require different relevant attributes and the criterion for effective instruction is the extent to which the medium used possesses the attributes appropriate to the task.

Salomon (1976) and Bates (1981) argue that attributes such as the level of technology or transmission have no effect on the learning process, and therefore cannot serve as a basis for a classification.

Salomon suggestes that media differ inherently from one other in the way in which the instructional message is translated into symbolic codes. Salomon (1979) makes a distinction between three symbolic systems (digital, analogic and iconic) and states that television is a very rich medium in that all three systems can be used. As the nature of the symbolic code determines the kind of mental skill used to process the information, it follows that only the translation of a message into different symbolic codes can result in different learning outcomes.

Changing the technical qualities of the media (for example, changing from a live lecture to a televised lecture) does not change the symbolic code of the message and therefore will not have any affect on the accomplishment of the learning task.

Symbol systems facilitate learning in one of three ways (Bates, 1981:84 quoted by Salomon, 1979):

- *by activating existing mental skills and providing practice in their use;*
- *by shortcutting difficult mental processes through representing knowledge in a new way;*
- *by supplanting or modelling the mental elaborations required i.e. demonstrating to learners how to move from point A to point B to incorporate new knowledge.*

According to this approach, the criteria by which the effectiveness of the medium should be judged are measures of the extent that such functions are accomplished. Salomon suggests that the accomplishments of these functions depends on the ability of the media to represent symbolically the nature of the learning task.

Bates agrees that ways in which instructional messages are translated into symbolic codes serve as criteria for evaluating the effectiveness of media. In his opinion, however, not enough is known about which mental skills are relevant to which learning tasks, which mental skills are best developed by television or the nature of the process of learning from television. Heidt (1980) adds that even though differences in learning outcomes cannot be attributed directly to differences in technology or transmission, these variables do have a considerable impact even if only on establishing the social context of the media. Criteria for evaluation should therefore be based upon these as well.

Another point that has received much attention from researchers of instructional media is the individual differences among learners. It is claimed that, in addition to the media attributes and the nature of the learning task, the learners' individual characteristics should be taken into consideration. Both Salomon and Bates discuss the different effects of media on learners of different levels of aptitude or skill and ability. Rushing (1982) found that students with distinct preferences for using one of the five basic senses while learning used different learning styles. Holliday et al (1977) reported that pictures and diagrams assisted students of low verbal ability. Palmer (1982) suggests that most research findings in the field of Aptitude Treatment Interaction (ATI) show that taking individual differences into consideration does not contribute to achievement. Winne (1977) found that ATI absorbed only 1-2 percent of the variance when teacher effectiveness was measured. Similar results were obtained by Wright (1977), Pascarella (1977) and others. Even Cronbach and Snow, who, in 1969 suggested that the interplay between learner characteristics and learning methods might offer more fruitful avenues for research, in 1977 stated that "lest the reader builds up false expectations, we should say at once that well substained findings regarding ATI are scarce" (p 6).

Merril (1985) attacks the ATI approach from a different point of view. He suggests that constructing programmes that cater for individual differences might render the learners system-dependent whereas the aim should be to make them independent of the system.

The Film Industry Approach

In an article entitled 'Educational Television: the gulf between researchers and producers', Hart (1982) presents the producers' point of view on evaluation focusing on ETV. Referring to articles written by researchers of educational television, he states "... (these articles) ... exasperated a number of producers who read them. I found them useful, if rather laboured examinations of what we are supposed to be doing. 'So what?' the producers said, 'If I didn't know that already, I wouldn't be employed'" (p 92). Hart proceeds: "I fear that ... much of the research into (Educational Television) is being done by academics who are blissfully unaware of the vast body of critical literature which has been around since the birth of the cinema" (p 95). The film industry is mainly profit-oriented and profits are only made through films for which the public is willing to pay. Although critics and the public evaluate films in terms of their artistic quality, many producers look at interest in the film and its ability to attract an audience as the only relevant criterion. Shary (1970) describes the various stages of the film production process at M.G.M. and emphasizes procedures for critical analysis employed throughout all the stages. According to him, M.G.M. employs 15 people who read about 25 000 books and stories each year and on the basis of their criticism, 30-40 stories are selected and presented for re-criticism to the company's producers. At the next stage the translation of the selected story into script form is evaluated by the producer himself, by heads of other departments and by censors and lawyers. Before the actual shooting of the film begins, it is evaluated again by heads of the relevant technical departments. Runs of the film as they are completed are evaluated by the whole production team as soon as possible. Evaluation plays a major role throughout the editing phase as well. The sneak previews serve as the last opportunity for inside criticism before the film is distributed and exposed to the public and the critics. Malkin (1979) says the criticism in the process of cinematic creations improves and refines the film at each stage of production. Although educational evaluators have not followed the same route as the film industry: this use of criticism, in many ways, parallels the use of formative evaluation in educational projects.

Indeed, many leading educationalists believe that the basic rationale for evaluation is that it provides information upon which rational decisions can be based (Cronbach, 1963; Scriven, 1972; Lewy, 1977). The most famous attempt to evaluate formatively an ETV-programme (or criticise in the terms of the criticism of film production) is the study on the production of "Sesame Street" (Ball and Bogats, 1970, 1971). It has been followed by many similar studies, the best known of which is probably the formative evaluation research of the programme Contact 3-2-1 (Mielke and Chen, 1983).

Film criticism is not guided by strict or well-defined research methodologies. Nevertheless, professional literature in the field abounds with articles and critical reviews analysing various films and the supposed causes of their success or failure. Articles are written by professional critics and famous producers whose films are recognised as great successes. There are also interviews with such producers. Vardac (1949), for example, has written about the difference between the characteristics of a play and a film, and considered the various aspects of adapting a stage play to cinematic form.

Bergonzi (1972) discussed the type of plots which have the potential to attract large audiences, and the differences between the situation of a novel and of a movie. Reisz and Miller (1953) considered a variety of editing techniques and the effect they have on creating the desired atmosphere. Cassirer (1970) was concerned with the philosophy of symbolic forms used in the cinema, and Durgant (1967) has written about the effect of films on audience feelings. Sylbert (1974), Kracauer (1947) and Eisenstein (1957) discussed extensively technical variables such as the angle of the shot, the screen composition or the lighting and their possible use in eliciting feelings such as suspense, relief of tension or tranquility from the audience. Some of the queries about producing attractive and interesting films have been examined by scholars involved in the production of ETV.

There is a large body of research about how variables such as the technical characteristics of a programme or its format affect the audience (Baggaley, 1980; Schramm, 1972; Geoffrey, 1978; Paluzzi, 1980). Coldevin (1976, 1980a, 1980b) summarizes the results of studies in the field and suggests considering three distinctive types of variables: technical variations, such as camera factors, setting, lighting, special effects, colour; content organization strategies such as opening and closing format, pacing and rhythm, repetition; and performer variables such as dress, age, sex, eye contact level etc. In his concluding remarks, Coldevin (1980a:95) states that the "chief culprit appears to be the low level of generalizability of the literature, stemming from the lack of replication of results from one population to another ... and within similar populations by other researchers".

The Status of Evaluation in the Field of ETV

Despite the rising status of evaluation in ETV and the great number of studies related to it, it seems that there are as yet no well-framed or widely-approved models for the evaluation of educational television. Furthermore, there is not even an agreement among scholars in the field as to the definition of 'a good educational television programme', its exact meaning, it structure or its component elements.

The complexity of the concept 'a good ETV-programme' implies that a great variety of measures are relevant when evaluation is considered, and indeed, in practice, different studies often employ a different set of evaluation criteria. The selection of criteria for a specific study is based mainly on the evaluator's intuition, experience or personal inclination. Intuition and experience enable almost any spectator or user of a particular object to make an evaluative comment about it and to declare whether he is appreciative of, or dislikes the object. The judgement of a casual spectator or even a connoisseur, however, is different from the judgement of a professional evaluator.

Eisner (1979) distinguishes between two kinds of judgements and says that the connoisseur reacts on the basis of his experience or intuition and is not expected to justify his judgement by a clear definition of his expectations or by using an approved set of measures. The professional evaluator on the other hand, strives to present conceptual justification for his evaluation. To do that he must define the desired features of the evaluated object, and thus specify the criteria upon which his judgement is based. These specifications must rely on explicit knowledge formulated in a sound theoretical framework.

Eisner's distinction is analogous to Haezrachi's (1954) who distinguished between the spectator and the critic's evaluation of a work of art. She writes: "Where the spectator is concerned with contemplating and enjoying ... the analytical critic ... analyses and evaluates the objects according to their appropriateness, their satisfaction, their relative importance, their originality and their being genuine". In this statement Haezrachi tries to provide a definition of 'a good work of art', though her criteria are very general.

The lack of full recognition of the usefulness of ETV evaluative studies may, at least, be attributed partly to the lack of professional status of ETV evaluation. The fact that evaluation studies are severely criticised by producers or directors whose intuition has developed through many years of active film-making, supports this notion.

It seems that if evaluation is to play a central role in ETV production, researchers and evaluators must attempt to formalize a theoretical framework from which a consensually-accepted definition of 'a good ETV-programme' can be derived. This definition can lead to the establishment of a set of criteria for making comparisons between findings of different studies, and thus serve as the basis for professional evaluation.

Defining the Content Universe of the Concept 'A Good ETV-Programme'

Defining the content universe of a multi-dimensional concept, and thereby identifying all the groups of items related to it, is the first step towards formulating a theoretical framework for the concept. Whenever a multi-dimensional concept is referred to, there should be a reliable way to decide whether a particular item is related to the concept or not. Criteria for evaluation should be derived only for items in the content universe of the concept, and a comprehensive evaluation study should use criteria in that content universe.

Mapping sentences were suggested by Guttman (1957) as instrumental in the process of constructing definitions for the content worlds of different multi-dimensional concepts, Shye (1978) comments on the usefulness of mapping sentences and their structure. He claims that a mapping sentence has two parts: the domain and the range. The domain includes at least one facet which differentiates between the various items according to set rules. In attitude studies, for example, researchers commonly refer to three different behavioural aspects: the affective, the cognitive and the instrumental aspects. If a researcher thinks that classifying attitudes by their behavioural aspects is important for the study, he will include in the domain of the mapping sentence one facet named 'behavioural aspect'. There will be three elements in the facet, namely affective, cognitive and instrumental. This strategy ensures that the researcher is aware that the attitude he is examining does not necessarily carry a unified meaning with respect to its behavioural aspects, and enables him to treat each of the three aspects differentially.

Each additional facet included in the domain is defined by an additional rule for the classification of the items, and thus increases the differentiation among them.

The content universe defined by the mapping sentence is made up of all the items from the domain facets. That is, if the domain has two facets (one with three elements and the other with four), the content world will have 12 distinctive items.

The range of the mapping sentence is made up of one facet which pre-defines the expected results of the measurement in terms of scales, for example, very high ... very low; very positive ... very negative, etc. The facets in the domain and the range are connected by non-formal sentences which describe relations between them.

Differentiating Among Items Associated with the Concept

Two rules for differentiating between items related to 'a good ETV-programme' have been drawn from the professional literature on the production and evaluation of ETV. These rules serve as the basis for the construction of mapping sentences which define the universe of content of the concept.

The overall quality of an ETV-programme is affected by both production variables (such as technical aspects, presentation form, presenter variables etc.) and by variables related to the content of the programme. Evaluating a programme on the basis of production variables would obviously not lead to the same conclusions as evaluating it by its content. Furthermore, a comprehensive picture of the quality can be achieved by the programme only if criteria of both kinds are used.

This implies that one rule of differentiation among items related to 'a good ETV-programme' would delineate which component of the programme is being evaluated, its content or its production variables. The facet differentiating the items would thus include two elements, namely content-related variables and production variables.

'A good ETV-programme' does not have exactly the same meaning for educational evaluators, media experts or people in the film industry. Criteria based either upon production variables or content might thus reflect different desirable characteristics of the programme if stated from each of the three relevant points of view. A programme may be consided excellent when evaluated from one point of view and at the same time be considered poor if another point of view serves as a basis for evaluation. For instance, a programme can be very attractive and employ all the unique attributes of television without teaching the audience anything. Alternatively, it can bring about cognitive gains even though it employs an unattractive and boring lecture form. Furthermore, the evaluation of a programme cannot be considered complete unless it employs criteria yielded by each of the different areas of study. It is therefore suggested that a second rule of differentiation among items associated with the concept 'a good ETV-programme' should delineate the point of view from which the programme is being evaluated. The facet differentiating items according to this rule would include three elements: the educational evaluation point of view, the instructional media point of view and the film industry point of view.

The Mapping Sentence

A mapping sentence defining the content world of the concept 'a good ETV-programme' is presented in Figure 1.

The mapping sentence shows that six distinctive types of criteria are relevant to the evaluation of ETV-programmes: criteria based on content and production variables reflecting the educational point of view (a1, b1 and a2, b1), criteria based on content and production variables derived from the instructional media point of view (a1, b2 and a2, b2) and criteria based on content and production variables representing the film production point of view (a1, b3 and a2, b3).

FIGURE 1: A MAPPING SENTENCE
defining the content world of 'a good ETV-programme'

An item belongs to the content world of the concept 'a good ETV-programme' if it is used to assess the programme by its

A. Programme Components *B. Point of View*

$$\left\{\begin{array}{l}\text{a1 Content}\\ \text{a2 Production Variables}\end{array}\right\}\ \text{from the point of view of}\ \left\{\begin{array}{l}\text{b1 Educational Evaluation}\\ \text{b2 Instructional Media}\\ \text{b3 Film Production}\end{array}\right\}$$

Achievement

$$\text{as}\ \left\{\begin{array}{c}\text{High}\\ .\\ .\\ .\\ .\\ \text{Low}\end{array}\right\}$$

The Applicability of the Model

An empirical study was used to test the applicability of the model. The study examined whether evaluation criteria used by prominent researchers in ETV could be classified according to the suggested model.

To create a list of such criteria, the book Proceedings of the Third International Conference on Experimental Research in Televised Instruction (Baggaley, 1980), was analysed. Going through each of the ten articles in the book, this author wrote down all statements which specified explicitly, or could be interpreted as implying, some evaluation criteria.

As a result of this process, 158 criteria were identified. As the book contain a widely disparate articles approaching ETV production and evaluation from various angles, the resulting criteria list can be considered quite comprehensive. However, the list is not fully representative of all criteria ever used for the evaluation of ETV-programmes. Most of the criteria on the list can be classified according to the two differentiation rules suggested in this article. There are only two groups of criteria which could not be classified according to one of the two rules. The first group includes criteria dealing with the holistic attributes of a programme (such a programme or its contribution to cognitive gains).

These criteria can be used to evaluate the overall quality of the programme and relate it to the overall combined effect of content. These criteria cannot, therefore, be classified according to the first differentiation rule. As each criterion reflects either the educational point of view, the instructional media point of view or the film industry point of view, it can be classified according to the second differentiation rule. It should be noted that criteria stated in such general terms have very little potential to pinpoint problematic areas in the programme or contribute to its refinement.

The second group of criteria, which could not be classified according to the two differentiation rules, includes criteria which refer to specific characteristics of the programme (such as the use of a male presenter, humorous inserts or dramatic format). The criteria cannot be used to specify why each of these characteristics is important, or in which way they make a programme better. These criteria can be classified according to the first rule as they are based either on content or on production variables. As their relationship to the quality of the programme is often only vaguely implied, however, they cannot be classified according to the second differentiation rule. The other criteria on the list yielded to classification according to both differentiation rules.

Examples of criteria from each of the six types are presented in Figure 2.

FIGURE 2: EXAMPLES OF CLASSIFIED CRITERIA

B = The Evaluation Perspective

A		Educational Evaluation (Important attitudinal or cognitive gains) (b1)	Instructional Media (Use of unique attributes of television) (b2)	Film Production (The attractiveness of the programme) (b3)
= The Evaluated	Based on Content Variables (a1)	- The content is syllabus orientated (p 10). - It brings about enthusiasm for learning (p 10). - The transmitted knowledge is appropriate to the level of previous knowledge (p 115). - The programme has a vocabulary control (p 115).	- The content is better understood if slow-motion is used (p 46). - The continuity of action is part of the learning task(p 68) - Media attributes are relevant to the particular learning task (p 47). - The content is appropriate to realise the non-translatibility axiom (p 54).	- The content is considered interesting by the target audience (p 3). - The topic appears on a high-priority list of the audience (p 12). - Dramatic content (p 118). - Employs definite theme (p 12)
Component	Based on Production Variables (a2)	- The use of social modelling. - Asymmetrical placing of still visuals (p 127). - High-value segments with low appeal follow high appeal bits (p 69). - Pause between review segments (p 69).	- Can the programme be understood using "picture only"? (p 9). - Visual modality of response (p 40). - Simultaneous use of visual-auditory stimuli (p 46).	- The format is preferred by the target audience (p 3). - The presentation style is preferred by the target audience. - Use of catchy and recognizable features as signature and title tune (p 19).

As all the criteria appearing in Figure 2 were elicited from the same book (Baggaley, 1980), it was decided to omit the names of authors of each criterion and cite only the page number on which it is mentioned in the book.

The point of view of educational evaluators is represented in the items concerned with the extent to which a programme responds to the identified needs of the clients and with outcomes in terms of cognitive or attitudinal gains. Among the criteria listed in the book, those pertaining to the category (a1, b1) concentrate on questions such as whether the programme is syllabus orientated, whether it brings about enthusiasm for learning or whether the transmitted knowledge is appropriate to the level of previous knowledge.

Evaluation by production variables from the educational frame of reference (a2, b1) is concerned with criteria such as the appropriate placement of high-value segments throughout the programme and pauses before review segments to enhance learning, etc.

Criteria which represent the views of media experts tend to be used to define 'a good ETV-programme' as a programme which successfully takes advantage of television's unique attributes. Evaluation criteria based on content stated by media experts (a1, b1) thus address mainly the question whether attributes of television are relevant to the particular content or learning task.

Evaluation based on production variables from this point of view (a1, b2) is concerned with the simultaneous use of visual-auditory stimuli, the visual modality of the presentation and with the use of special techniques such as slow-motion.

The criteria which represent the point of view of people from the film industry can be used to examine a programme mainly on its ability to attract and entertain the target audience. From this point of view the content would have to be interesting and enjoy competitive appeal (a1, b3). The production variables would also have to match target audience preferences (a2, b3).

Conclusion

The relative importance of production or content-related criteria, reflecting the three different points of view, may vary according to the researcher's approach or the study objectives. It could also be influenced by variables such as the stage of production at which the evaluation study is implemented, the nature of the target audience and factors such as the intended use of the programme (for broadcasting or non-broadcasting) or the intention of the programme (syllabus-oriented or transmission of general knowledge). Nevertheless, it seems that, in all cases, only a programme which is satisfactory with respect to the whole variety of criteria can be considered a really 'good ETV-programme'. No overall evaluation can, therefore, be complete unless criteria representing all three points of view are used systematically.

If, as a result of time or budget limitations, an evaluator does not intend to use measures relating to all the criteria, he should be aware that he will not evaluate all the relevant aspects of the programme. In case he decides to focus on a particular aspect and ignore the others, he should recognise that this one aspect does not reflect the overall quality of the programme and this fact should be emphasized when he summarizes his findings.

An evaluator who selects criteria for evaluating without being aware of the different perspectives of ETV-programmes may base his conclusions upon a random collection which neither displays fully the relevant advantages or disadvantages of the programme nor highlights those areas in which improvement is greatly needed.

Once there is a comprehensive definition of the concept 'a good ETV-programme', further research should explore the empirical relations between the different perspectives of the concept and attempt to verify the assumed relations between each predictive criterion and the goals it is meant to predict.

Gathering research evidence in those areas will contribute to the formulation of a theoretical framework for the evaluation of ETV-programmes. This theoretical framework will provide the conceptual justification for the selection of a set of criteria upon which professional evaluation of ETV can be based.

REFERENCES

Aversa, F and Morrison, G R 1979. Toward more precise definition and evaluation of televised educational programmes. Educational Technology, August, 1978, 5-11.

Baggaley, J (ed.) 1980. Proceedings of the Third International Conference on Experimental Research in Televised Instruction. Memorial University of Newfoundland. August, 1980.

Ball, S and Bogatz, G A 1970. The First Year of Sesame Street: An Evaluation. Princeton, Educational Testing Service.

Ball, S and Bogatz, G A 1971. The Second Year of Sesame Street: A Continuing Evaluation. Princeton, Educational Testing Service.

Bates, A W 1981. Some unique educational characteristics of television and some implications for teaching and learning. Journal of Educational TV and other Media, Vol. 7, No. 3, 79-86.

Bergonzi, B 1972. The Situation of the Novel. Middlesex, England, Penguin Books.

Bloom, B S et al 1971. Handbook on Formative and Summative Evaluation of Students Learning. New York, McGraw-Hill.

Boucher, B G, Gottelieb, M J and Morganlander, M L 1971. Handbook and Catalog for Instructional Media Selection. Engelwood Cliffs.

Bretz, R 1971. A Taxonomy of Communication Media. Engelwood Cliffs.

Cassirer, E 1970. The Philosophy of Symbolic Forms. Yale University Press.

Chu, G C and Schramm W 1968. Learning from Television: What the Research Says. Stanford, California.

Coldevin, G O 1976. Comparative effectiveness of TV production variables. Journal of Educational Television and other Media, Vol. 2, No. 3, 87-92.

Coldevin, G O 1980. Formative research in television presentation strategies: guidelines for effective production. J Baggaley (ed.), Proceedings of the Third International Conference on Experimental Research in Televised Instruction, Memorial University of Newfoundland, 62-80.

Coldevin, G O 1980. Experimental research in television production techniques and presentation strategies: Current direction. Journal of Education Television and other Media, Vol. 6, No. 3, 92-97.

Creaser, H 1981. The impact of media on learning. Journal of Educational Television and other Media, Vol. 7, No. 3, 99-101.

Cronbach, L J 1963. Course improvement through evaluation. Teachers College Record, Vol. 64, 672-683.

Cronbach, L J and Snow, R E 1969. Individual Differences in Learning Ability as a Function of Instructional Variables. Stanford University.

Cronbach, L J and Snow, R E 1977. Aptitudes and Instructional Methods. New York, Irvington.

Duncan, C J 1969. A survey of audiovisual equipment and methods. D Unwin (ed.), Media and Methods: Instructional Technology in Higher Education. Maidenhead.

Eisenstein, S 1957. Film Form, The Film Sense, Meridian Books.

Eisner, E 1979. The Educational Imagination on the Design and Evaluation of School Programmes. New York, MacMillan.

Durgant, R 1967. Films and Feelings. London, Faber and Faber.

Eraut, M, Good, L and Smith G 1975. The Analysis of Curriculum Materials. Brighton, University of Sussex, Educational Area, Occasional Paper 2.

Geoffrey, R 1978. The effectiveness of colour in learning from educational television. Journal of Educational Television and other media, Vol. 4, No. 1, 21-23.

Guttman, L 1957. Introduction to facet design and analysis. Proceedings of the Fifthteenth Internationl Congress of Psychology. Brussels, Amsterdam, North Holland Publishing.

Haezrachi, P 1957. The Contemplative Activity. London, Allen and Unwin.

Hart, L 1982. Educational TV: The gulf between researchers and producers. Journal of Educational TV and other Media, Vol. 8, No. 2.

Heidt, E U 1980. Technologies and symbol systems: analysing differences between and within media. J Baggaley (ed.), Proceedings of the Third International Conference on Experimental Research in Televised Instruction. Memorial University of Newfoundland. August, 1980, 36-61.

Holliday, W G, Brunner, L L and Donais, E L 1977. Differential cognitive and affective Responses to flow diagrams in science. Journal of Research in Science Teaching, 14(2), 129-138.

Kemp, J E 1971. Which medium? Audiovisual Instruction.

Kracauer, S 1947. From Caligari to Hitler, a Psychological History of the German Film. Dennis Dobson, Princeton University Press.

Leech, D 1980. What makes television the right medium? Journal of Educational TV and other Media, Vol. 6, No. 1, 25-30.

Lesser, G S 1974. Children and Television: Lessons from Sesame Street. New York, Random House.

Malkin, Y 1979. The Function of Criticism in the Process of Cinematic Creation. Theses submitted for the degree of Phd at Tel-Aviv University, Israel.

McNaught, P C 1979. Educational television, 1969-1979. Journal of Educational Television and other Media, Vol. 5, No. 1, 4-5.

Merril, M D 1975. Learner control: aptitude treatment interaction. AV Comunication Review, 23(2), 217-225.

Mielke, H W and Chen, M 1983. Formative research for 3-2-1 contact: methods and insights. J A Howe (ed.), Learning from Television Psychological and Educational Research. London, 31-55.

Nevo, D 1983. The conceptualization of education evaluation: an analytical review of the literature. Review of Educational Research, Vol. 53, No. 1, 117-128.

Palmer, C R 1979. Trends in research: aptitude treatment interaction. Journal of Educational TV and other Media, Vol. 5, No. 1, 14-15.

Paluzzi, J V 1980. Interest and presentational formats. J Baggaley (ed.), *Proceedings in the Third International Conference on Experimental Research in Televised Instruction*. Memorial University of Newfoundland, 22-35.

Parlett, M 1974. New evaluation. *Trends in Education*, No. 34, 12-18.

Pascaralla, E T 1977. Student motivation as a differential predictor of course outcomes in personalised system of instruction and conventional instructional methods. *Journal of Educational Research*, 71(1), 21-26

Provus, M M 1971. *Discrepancy Evaluation*. Berkely, California, McCutchan.

Reisj, K and Millar, G 1953. *The Technique of Film Editing*. London, The Focal Press and New York.

Rushing, B 1982. Cognitive style and videotype. *NSPI Journal*, October, 36-39.

Salomon, G 1976. A cognitive approach to media. *Educational Technology*, May, 25-28.

Salomon, G 1979. *Interaction of Media, Cognition and Learning*. San Francisco.

Schramm, W (ed.) 1972. *Quality in Instructional Television*. The University Press of Hawaii.

Scriven, M 1972. Prose and cons about goal-free evaluation. *Evaluation Comment*, No. 3, 1-4.

Shary, D 1970. As told to Charles Palmer. *A Case History of a Movie*. New York, Random House.

Shye, S 1979. On the search for laws in the behaviour sciences. S Shye (ed.), Theory Construction and Data Analysis in the Behavioural Sciences. San Francisco, Jossey-Bass.

Stake, R E 1981. The countenance of educational evaluation. Evaluation News, Vol. 2, No. 2, 148-152.

Stephen, K D 1975. A matrix for quality. Journal of Educational TV and other Media, November, 2-6.

Stuffelbeam, D L 1972. The relevance of the CIPP evaluation model for educational accountability. SRIS Quarterly 5, 3-6.

Sylbert, P 1974. Final Cut, The Making and Breaking of a Film, a Continuum Book. New York, The Seabury Press.

Totsi, D T and Ball, J R 1969. A behavioural approach to instructional design and media selection. AV Communication Review, Vol. 1.

Tyler, R W 1950. Basic Principles of Curriculum and Instruction. Chicago, University of Chicago Press.

Vardac, A 1949. Stage to Screen, Theatrical Method from Garrick to Griffith. Cambridge, Harvard University Press.

Williams, F and Monty, C S 1976. ITV/ETV evaluation: What we should expect. Educational Technology, May, 35-38.

Winne, P H 1977. Aptitude treatment interaction in an experiment on teacher effectiveness. American Educational Research Journal, 14(4), 389-409.

Wright, D P 1977. Interaction between instructional methods and styles of concept learning. Journal of Educational Research, 70(3), 150-156.

SOURCES OF EVALUATION CRITERIA IN EDUCATION, WITH PARTICULAR REFERENCE TO THE EVALUATION OF ETV

A Lewy

The Nature of Evaluation Criteria

One encounters evaluation activities in a very broad range of situations. Evaluation may occur as a spontaneous reaction to a certain stimulus. One may frequently show a liking or disliking for something, such as food, odour, landscape, music or a work of art, without being able to specify one's reasons for such reactions. The reasons for liking or disliking something may stem from sensuous experiences of a subjective or less subjective nature. One may like a certain type of food because its taste is associated with childhood memories or because it is spicy. Reasons for liking or disliking something may also be rooted in highly abstract conceptions about what one regards as meritorious. An example of this is liking a short story because it contains a moving description of a tragic event.

Evaluation activities are not necessarily spontaneous reactions to stimuli. On the contrary, they are most frequently premeditated and carefully planned processes which are carried out to provide judgemental statements about the worth of objects being examined. Examples of such judgements are the wine taster being asked to determine the quality of the vintage, the gourmet preparing a review of a restaurant for the local magazine, the critic writing an essay on a work of art, a research officer comparing the merits of two competing educational programmes. All are engaged in evaluation activities.

Evaluation activities may differ from one another in several ways: they may be either spontaneous or premeditated; refer to sensuous experiences or to highly abstract concepts. They may rely on fully subjective or inter-personally endorsed justifications, and result in a publically announced value statement or a private conviction. Nevertheless, in all cases the object being examined possesses a certain attribute or series of attributes which form a basis for the judgement. Any alteration to these attributes may alter the judgement. The attributes taken into consideration in making evaluative judgements about a particular object are referred to as criteria.

This article aims to identify some characteristic features of criteria used to evaluate educational television programmes.

The subsequent sections deal with the following four major topics:

* *The meaning of criteria and standards, and how they are determined and used in the context of educational evaluation.*

* *Evaluation criteria in literature and art, and their relevance to evaluation criteria in education.*

* *Checklists and inventories as evaluation criteria. This section deals with evaluation criteria in education in general and serves as a background for dealing with a particular type of educational programme, the ETV programme.*

* *Criteria for evaluating ETV programmes.*

The reader will notice that the article starts with issues related to educational evaluation in general. It examines similar patterns of evaluation activities across two different disciplinary areas: literature and education. It also provides a checklist for evaluation in general, and for ETV in particular.

The structure reflects the view that evaluation of each particular educational medium should be related to evaluation in education in general, and to evaluation activities across a broad variety of disciplinary areas.

Criteria and Standards in Evaluation Studies

The term criterion in contemporary philosophy denotes something which can show the truth or falsity of a certain state of affairs. According to Wittgenstein, a criterion is "something by which one may be justified in saying that the thing is so, and by whose absence one may be justified in saying that the thing is not so, all by its logical necessity" (Kenny, 1967: 2;258). For example, a criterion for 'angina' is a certain type of bacillus. Alternatively, the term criterion has been used as a scale by which something can be judged or tested. Both meanings of this term are retained in contemporary research studies of various types, mostly referring to the variables of the study.

In a more restrictive sense, criterion is sometimes used to denote dependent variables by which the effectiveness of a particular treatment is examined. Thus, for example, one may encounter experimental studies for examining the worth or merit of an ETV programme in which a specified skill or type of information serves as a criterion variable. The researcher 'tests' whether the skill or the knowledge was acquired, and on the basis of the observed results, he adopts a stance on the worth of the treatment. A semantic distinction should be made between the words 'standard' and 'criterion'. As Medley (1985:1 735) states "a criterion is an aspect or a dimension of the quality to be evaluated, which is to be assessed and then compared with an arbitrary standard or level of this quality as a basis for evaluating it". In other words 'criterion' refers to a significant aspect of a thing being examined, whereas 'standard' denotes a certain cut-off point on the continuum of a measure representing the level of competence.

One may set a standard of 80 percent correct response to test a particular skill. In this case, the attainment of such a standard on the specified criterion serves as evidence of the worth of the programme evaluated.

In some contexts 'standard' refers to an authorative model which can be used as a basis for comparison to determine the quantity, excellence, and correctness of other things. The Oxford Dictionary defines 'standard' as a definite degree of any quality, viewed as a measure of what is adequate for some purpose. This definition implies that standards used in practice may vary from one case to another: an adequate degree of a quality in one case may be less than adequate in another case. One may establish a certain cut-off point on a test for a group of gifted children viewing a particular ETV programme, and a different point for a group of average children viewing the same programme.

Selecting criteria and setting standards are complementary activities, and in carrying out evaluation studies both require considerable attention. Nevertheless, in some areas of research, the distinction between the two is not strictly observed. Thus, for example, the American Educational Research Association (AERA, 1955) set up a series of criteria for judging the adequacy of achievement tests. These guidelines were published under the title 'Technical Recommendations for Achievement Tests'. A later edition of a revised set of recommendations carried the title "Standards of Educational and Psychological Tests and Manuals" (APA 1966). A team led by Stufflebeam published a book entitled "Standards for Evaluation of Educational Programmes, Projects and Materials" (Joint Committee, 1981).

In both cases the term 'standards' is used for the word 'criterion' as defined above. This imprecise use of terms could confuse the planner of a evaluation study. Planning an evaluation study entails both selecting criteria and setting standards.

The implications of the distinction between criteria and standards are thoroughly documented in articles dealing with methodological aspects of evaluation: Scriven (1967) addresses himself to problems of selecting criteria and Glass (1978) discusses problems of setting standards.

Scriven (1967) lists evaluation questions which should be answered on the basis of measurements, related to specific criterion variables. These questions are:

- *How well does a particular instrument or treatment perform (with respect to specific criteria)?*
- *Does it perform better than another instrument (also in terms of specified criteria)?, and*
- *What does a particular instrument do? i.e. what variables from the group we are interested in are significantly affected by its application?*

Scriven (1967) states that a series of criteria (and not a single criterion) should be used to answer each of the above questions. By this, he alludes to the importance of considering multiple outcomes for evaluating the worth of a particular instrument or treatment. Scriven does not address the question of selecting criteria. He is satisfied with writing about "criteria in which we are interested", i.e. the evaluator has to deal with criteria which constitute the focus of interest. Does Scriven specify to whom the pronoun "we" refers? It may refer to the evaluator, to the sponsor of the evaluation study, to the user of the instrument, to the recipient of the treatment, or to members of the community in general. Nevertheless, Scriven does not view the criteria as an arbitrary selection of the measurable aspects of an instrument or treatment.

The selection of criteria constitutes an inherent part of the evaluation activity, and decision-makers are obliged to justify both their selection of the criteria and the assignment of differential weights to them. To meet such demands a broad variety of approaches to justify the selection and weighting of evaluation criteria has been devised.

The question of performance standards is examined by Glass (1978). He lists several attempts by measurement experts to deal with this problem. For example, he quotes Mager (1962:44):

> *If we can specify at least the minimum acceptable performance standard against which to test our instructional programmes, we will have a means to determine whether our programmes are successful in achieving our instructional intent. What we must try to do, then, is to indicate in our statement of objectives, what the acceptable level of performance will be, by adding words to describe the criterion of success.*

Mager provides an example of an instructional objective and its associated standard: "The student must be able to solve at least seven simple linear equations within a period of thirty minutes".

Ralph Tyler (1973:105) specifies performance standards in the following words:

> *The child would be presented with a large enough sample of examples to furnish reliable evidence that he could recognize the letters of the alphabet.*
>
> *He could associate the sounds with each letter, alone and in words, and he could recognize the one hundred most common words. A child has demonstrated mastery of a specified knowledge, ability or skill when he performs correctly 85 percent of the time. (Some small allowance, like 15 percent, is needed for lapses common to all people).*

Glass (1978:238) rejects these attempts at setting arbitrary performance standards, and views them as "pseudo-quantification, a meaningless application of numbers to a question not prepared for quantitative analysis. A teacher or a psychologist cannot set meaningful standards of performance as imprecisely defined as correctly spelling words called out during an examination period". While Glass (1978) does not suggest specific ways of setting standards, he fully accepts the conceptual distinction between selecting criteria and setting standards, and recognises the importance of both activities in an evaluation study.

In this paper we shall limit our interest to issues related to evaluation criteria, i.e. variables, traits, phenomena, occurences, etc. associated with an object being evaluated. Problems related to setting standards, i.e. establishing the minimal acceptable level (degree of scale value) of the measured quality will remain outside the scope of this paper.

Formal Categories of Evaluation Criteria

Evaluation experts tend to distinguish between groups of evaluation criteria on the basis of their formal characteristics. Scriven (1967) distinguishes between intrinsic and pay-off evaluative criteria. Intrinsic criteria are concerned, inter alia, with evaluating programme goals. According to Scriven this evaluation cannot be done on the basis of pure empiricism. Another intrinsic evaluation task is consistency analysis in which divergences between (a) espoused, (b) implicit, and (c) tested goals are examined. In contrast to the intrinsic evaluation criteria, the 'pay-off criteria' examine programme effects. In a pay-off evaluation study the programme being evaluated is conceived as an independent variable, and its effects are conceived as a series of dependent variables.

Another distinction between formal categories of evluation criteria is the universal and particularistic dichotomy (Eden, 1984).

Criteria derived from broad theories of curricula or teaching-learning are frequently conceived as universally valid, and are seen to be adequate to evaluate any educational programme. Examples of universal criteria are: comprehensibility of verbal and non-verbal symbolic messages in instructional materials, the learners' interest in the theme of the lesson, correctness of facts conveyed, etc.

Shepherd (1972) suggested examining sentences to determine the proportion of questions to instructions (e.g. read, write, discuss, handle equipment) as a criterion for evaluating textbooks. Such criteria are frequently labelled external criteria as their source is located outside the programme being examined. By contrast, particularistic or internal evaluation criteria are derived from programme documents, and may be unique to a particular programme.

Adar and Fox (1978) provided an interesting example of internal evaluation criteria in their analysis of a history textbook and supplementary curriculum materials used in the higher grades of the Israeli primary school. They provided criteria such as: instances of diverse view about a particular social or value issue; a balanced presentation of views of two parties in conflict with each other; encouraging the reader to take a stance on a controversial issue; etc.

A third distinction between formal categories of criteria is the 'procedural' versus 'substantive' distinction (Kaiser, 1983). Procedural criteria deal with the adequacy of the evaluation process. Examples of such criteria are:

* the participation of interested parties in judging the worth of the programme, such as women's organizations, minority groups, representatives with relevant conflicting social values, etc.;

* carrying out the negotiation and decision-making processes according to formally endorsed rules;

complying with the law, etc.

Some evaluators refrain from stating substantive evaluation criteria which require the specification of particular values. They content themselves with monitoring the evaluation activities in a way which may ensure the observance of procedural criteria. Other evaluators focus on specifying substantive criteria, i.e. specific values derived from social and educational theories or from the documents of the programme being evaluated.

Evaluation Criteria in Literature and Art

Instructional materials are cognate to literature. Both contain verbal materials and are published in the form of books: the parallels have led to attempts to evaluate curriculum materials in the mode of literary criticism. Both curriculum evaluators and literary critics have emphasised the affinity between the evaluation work on curriculum materials and literary criticism. Willis (1978), a curriculum expert, writes that curriculum:

> *... is an art as well as science. While it is well for the curriculum worker to know the scientific principles, the empirical data and the technological applications that inform much of his practice, it is equally well for him to be conversant with those aspects of arts which come into play when what happens with a curriculum or in a classroom inevitably spills out beyond those scientific constructs he has chosen to employ (p 93).*

Similarly Hirsch (1976), whose writings are well known in literary criticism, emphasises the common nature of cultural objects based on the utilization of the language:

> *A technical essay, an ordinary conversation or a poem has, therefore, necessary, inherent values; of course, the values are different, but the structure of arguments for their existence is the same. It follows that there is no sound reason for isolating literature and art in mysterious ontic realms apart from other cultural realities (p 109).*

Literary criticism, as a field of disciplined inquiry, has a longer history than curriculum evaluation and its principles are characterized by a high level of applicability to a wide range of cultural phenomena. An incipient field of enquiry such as curriculum evaluation may, therefore, benefit from examining problems and issues related to the topic: evaluation criteria in literature.

Selected topics from literary criticism will be mentioned here, and their relevance to the evaluation of instructional material will be examined.

Consensus as a Basis for the Validity of a Criterion

Literary critics do not have checklists and the critic's work and his critique are much more idiosyncratic than that of the curriculum evaluator. Indeed, in literary criticism a great variety of alternative criteria are used. Hirsch (1969) queried whether there were privileged criteria in literary evaluation, and whether there was any possibility of demonstrating the superiority of one evaluation mode to another.

This question is also asked by curriculum evaluators (Rossi and Freeman, 1982; Rieken and Boruch, 1974). Interestingly enough, the response to this question by literary critics is also dominant in the field of curriculum evaluation. Hirsch (1969) makes reference to David Hume's view that one may recommend criteria which are commonly accepted by experts or by the intellectual aristocracy.

Hume observed a high level of consensus among well-educated people in judging the worth of a literary work, and hence expressed confidence in the validity of such judgements.

In curriculum evaluation, consensus among experts is considered a basis for judging the importance or relevance of a particular evaluation criterion.

Hume's pragmatic canon rests on social agreement. In addition to this, there are frequent references in the contemporary sciences to the authority of an 'invisible college'. This is a group of experts who are viewed as leaders of a well-defined domain or a disciplined enquiry. Indeed, authority has always been an important aspect in the judgement of literature and curriculum materials. Hirsch (1969) noted however that

> *sheer authority has always played a role in literary judgement, but the institutions in which it has resided have become constantly more weak and diffuse. Aristotle's rules gave way to the Humean consensus, which in turn gave way to the modern fragmentation of all traditional jurisdictions. Yet anyone who questions traditional authority is thrown back upon his standards of taste, just as Luther, in questioning papal authority, was thrown back upon his own inward judgement in matters of faith. For some centuries now we have all been literary protestants without Pope or priesthood (p 23).*

One may add that multiplicity of competing authorities and "protestants without Pope and priesthood" characterises the fields of both educational evaluation and literary criticism.

Intent and Goal-Free Evaluation

Another pervasive issue in literary criticism is the question of intent.

To what extent should the author's intent serve as a basis for establishing evaluation criteria? Should the critic question the aim of the work, and whether it has been realised? The 'intentionalists' claim that these are the only legitimate questions for judging literature. Several objections have been raised to this view. In the first place one frequently hears that it is impossible to know the aim of a work. Secondly, what difference would the success of the aim make if one does not value the aim.

Hirsch (1969) lists arguments to refute these objections. Replying to the first objection, he points out that though one can never be certain of the aim of a work, one might still reach an acceptable level of certainty on the aim. As to the second argument, he feels one may substitute the question, "What aim is the work trying to achieve?" with "What should constitute an appropriate aim for such a work?"

The importance of knowing the author's intention is also discussed by Hospers (1969). He says the author's testimony is not necessarily the most valid indication of the truth implied by the text. The content of a particular statement and the circumstances under which it was uttered may help the reader determine the real intention of the writer or speaker.

The sentence "I went to the pictures last Tuesday" may have different meaning depending on the circumstances of its utterance. It may imply that the speaker really went to the pictures, or he believes that he did so, or that he has intentionally lied to mislead somebody. It may also serve as an example for a grammatical rule, without saying anything about the speaker's actual behaviour. The author's explicit statement about his intention may be useful for interpreting the meaning of a text, but it does not necessarily reveal what is transmitted by the text. Observing which passages of a novel contain great passion and intensity, which themes are reiterated, and which characters are treated with sympathy may provide 'belief-cues' for making inferences about the truth of the text.

Hospers (1969) attributes greater significance to intentions and implications embodied in a work and inferred from it, than to those explicitly stated by the author.

Art critics present additional arguments against taking the manifestly declared intentions of the author as a basis for judging the worth of a work of art. Harrison (1968) claims that the critic does not use readily available criteria for making judgements. The critic is the source of new ideas. In this respect the critic acts differently from a livestock judge. The judge of livestock may learn new principles such as 'a deep chest is a sign of a good chest' and eventually learn all the rules relevant to judging livestock. The art critic never learns all the rules. The judge of livestock has a pattern for establishing criteria and worth. By contrast the art critic does not operate within a closed set of criteria. In aesthetics there is no ultimate definition of correct judgement. There is always room for new books about Hamlet.

Henze (1968) believes that a series of criteria for judging a work of art can never be complete. George Boas's historical analysis of the Mona Lisa shows that, on the basis of Vasari's experiences, one could not predict what Freud and Pater's experience of the picture would be.

The intentions of a particular set of curriculum materials are of concern to curriculum evaluators. Several curriculum evaluation models give attention to the relationship between intent and its realisation. Tyler's (1949) model of evaluation is based on examining whether specified instructional objectives are attained. Stake's (1967) countenance model expands the domain of intentions to be examined within the evaluation framework of the study. It suggests that not only intended outcomes, but also intended antecendents and programme transaction should be evaluated. Thus for example, one has to ascertain that the actual conditions for using a particular programme (in terms of the pre-requisite knowledge of learners, the skills of teachers, etc.) correspond with the author's intention.

Many curriculum checklists contain questions about the specification of instructional objectives, guidelines for using the materials (i.e. intentions concerning classroom transactions), detailed descriptions of the target population and the conditions in terms of space, staff, equipment, etc.

These models adopted an 'intentionalist' approach to evaluating a particular object. According to this approach, evaluation starts with gathering information about intentions and continues with finding out whether the intention has been realised. In contrast to this pre-ordinate pattern of evaluation activity is the contemporary practice of starting with the text. Evaluation criteria are derived from the whole package of instructional materials and the actual teaching/learning situation.

This means one has to identify goals that can be attained and pursued by using these materials. Are these goals valuable? Is there a chance that they will be attained?

This approach is represented in the phrase 'curriculum potential' coined by Ben-Peretz (1985). She defines this as the totality of learning experience. It can be derived from a particular set of resources, such as textbooks, worksheets and audio-visual aids. Ben-Peretz illustrates this concept with an example:

> *A passage in biology curriculum materials concerning water composition by plants may originally have been intended to serve as an introduction to scientific inquiry. According to the developer's specification students may be asked to discuss the specific knowledge it imparts. When reinterpreted, the passage may be viewed as illuminating an interaction between science and society and the relevance of applied science to everyday life.*

> *It can be viewed as a vehicle to impart a positive attitude towards research and identification with scientists. The planned introduction to an instance of inquiry can thus be transformed into a series of learning activities and experiences that were not planned by the developers.*
>
> *These activities constitute a realisation of some potential inherent in the materials. The choice of curriculum potential to be realised in a particular situation is usually determined by perceived student needs and teaching preferences (2;1 247).*

The idea that a series of curriculum materials should be evaluated not only on the basis of the attainment of intended goals but also on the basis of considering the attainment of unintended outcomes, resulted in the notion of Goal-Free Evaluation (GFE). The Goal-Free Evaluator believes that he can bypass the producer's formulation of goals because the crucial question is not what the producer intended to do, but what was actually achieved (Scriven, 1972). Thus, rather than paying attention to goals specified by the producer, one should try to reconstruct them through observing the actual practice.

Intrinsic and Extrinsic Criteria

The distinction between intrinsic and extrinsic criteria appears both in literary and art criticism. Hirsch (1969) claims that in the history of criticism four modes of literary evaluation have envolved. One of these modes is based on extrinsic, and the rest on intrinsic considerations.

Evaluating a literary work in terms of its relationship to external factors is an example of using extrinsic criteria. A proponent of this view is Plato, who states that literary work should be judged good only if it is good for the state. Tolstoy is another outstanding representative of the extrinsic critics.

On the other hand, the intrinsic critics claim that a work should be judged good or bad based on the extent to which it fulfills some intrinsic imperatives pertinent to its genre. Critics have different definitions of the scope of art-genre. The adherents of the broad-genre theory favour the grouping of all literature under a single common theory. Other adherents of this theory (mainly those who pursue the discipline of aesthetics) are willing to include all branches of art into a single theoretical framework. A more restricted definition of genre is applicable to a specific kind of literature, such as comic-novel or lyric-poem. The adherents of this definition claim that "the right way of judging a comic-novel, for example, is not according to universal or external literary criteria, but according to the criteria specifically appropriate to comic-novels" (Hirsch, 1969;26). Finally, a third group of intrinsic critics have adopted an individualistic approach. They claim each work should be judged according to intrinsic imperatives inherent to that particular work. Each work is unique and cannot be included even in a narrowly defined genre.

The distinction between extrinsic and intrinsic criteria is also manifest in curriculum evaluation. But in the field of curriculum evaluation traditionally, the use of extrinsic or pay-off criteria has been dominant.

The growth in the use of intrinsic criteria is a recent development, but as Eraut (1985) states, the development of techniques for intrinsic evaluation is still in its infancy. Eraut lists variants of intrinsic evaluation methods, such as content analysis, structure and sequence analysis, consistency analysis, etc. Mann (1969) introduced an additional variant of intrinsic analysis which he called curriculum criticism. He argued that the purpose of such analysis is disclosing meaning and relating materials to the social context.

The distinction between individualistic and genre-related intrinsic criteria has also appeared in curriculum evaluation. Eden (1984) refers to these categories as external and internal evaluation criteria. Consistency between stated instructional objectives and suggested learning experiences may constitute an example of external evaluation criteria. External evaluation criteria can be used universally for all kinds of instructional materials, or at least for those genre that contain explicity stated instructional materials.

Internal criteria also refer to those criteria inherent in a particular curricular object (textbook, worksheet and audio-visual programme), and have been derived through a systematic analysis of the object itself.

Checklists and Inventories of Evaluation Criteria

Several recently published evaluation textbooks contain a large assortment of checklists recommended as instruments for programme evaluation. For example, Woodbury's (1979) book, designed to meet the needs of library acquisition departments, contains a compilation of checklists released in the last 30 years. The author does not examine the rationale behind particular checklists, or attempt to compare them analytically with one another, or judge their quality and practicality.

The reader may be of the opinion that the use of a checklist for evaluating instructional products is a methodological innovation generated by the curriculum movement of the late 1950's. A thorough review of the literature reveals, however, that as early as the 1930's checklists were generated and used for evaluating 'educational talking pictures' (Cambre, 1982). Devereux (1933) described conferences of the Electrical Research Products Inc. (ERPI) which judged each element of an educational film. He found that research and scientific conferences played a major role in shaping the development of the educational 'talking picture' from its start, and described the way group judgements, based on objective considerations and approved standards, were used in this context.

Group conferences produced checklists. The use of these checklists was supposed to orient producers to developing 'errorless' educational films and help users select films to suit their needs. The ERPI checklist contains items related to the objectives of the picture, its content, the technical-audio-visual elements, the integration of the film with other curriculum materials and the general effectiveness of the programme.

Goodman (1941) later described phases in the development of such checklists. He suggested starting with a survey of literature and the study of procedures for using films in schools to examine the worth of the producers. These phases would be used to prepare preliminary checklists which would be reviewed by committees. The final phase would be to design small experiments to evaluate the suggested criteria.

Glass (1969) claims that evaluation checklists were first used for the accreditation activities of organizations such as the National Council for the Accreditation of Teacher Education (NCATE)(Wheeler, 1980) and the National Study of Seconday School Evaluation, (1960). A growth in the quantity of evaluative checklists for programme evaluation followed these activities.

Eraut (1975) lists four reasons for the burgeoning of curriculum evaluation checklists. Firstly, one encounters a great variety of curriculum development patterns: a particular checklist which suits one development pattern does not necessarily suit another. Secondly, there is no unanimous acceptance by education experts of a single learning theory. Thirdly, the context in which the programme is used changes from case to case. Finally, learners differ from one other.

Eraut's remarks suggest that one may construct a classification scheme of evaluation situations using the following dimensions:

* pattern of programme development,
* learning theory adopted,
* the context in which the programme is used, and
* the characteristics of the user.

The items in a checklist should be constructed or selected on the basis of the evaluation situation using the above-described four-dimensional scheme. For example, the checklist for evaluating an educational programme based on the principles of 'learning by discovery' is likely to be entirely different from that for evaluating a school-based curriculum project.

While it may be unrealistic to expect the number of checklists to match the number of potential evaluation situations, there is still a possibility of selecting relevant items from a larger pool of items (Choppin, 1978). One could also employ an ecletic approach (the 'cafeteria approach') for compiling measurement instruments. The cafeteria approach for selecting evaluation criteria was established, and has been successfully used for evaluating courses in higher education. Its operation is described by Braskamp et al (1984). The system consists of a pool of labelled and classified criteria.

An institute may either select a uniform subset of items for evaluating a particular group of courses from the criteria, or permit each teacher to select those criteria which he considers adequate. Alternatively, the institute may combine the two above-mentioned possibilities by prescribing a set of uniform core criteria across all courses which should be complemented by each teacher according to his individual taste and inclination.

Programme evaluation checklists lack such systematization and consequently the evaluator has to adopt an available instrument or devise his own checklist.

Instances of Evaluation Checklists

Because of the lack of a consensually accepted classification scheme of evaluation criteria, it seems impossible to provide an exhaustive list of all types of checklists. Instead, one has to be content with an occasional set of checklists and evaluation forms which represents a broad variety of frequently used instruments.

In this paper preference is given to lists and instruments which are relatively new, i.e. those published in the last two decades. Most of the checklists used below were designed for evaluating instructional materials of a conventional type, such as textbooks, study kits, etc.

The reader interested in evaluating materials produced by, or for, advanced technology may profit directly and indirectly from examining these forms. In a direct way, the criteria used in these instruments may be relevant to the reader's own needs. In an indirect way, he may uncover the underlying rationale of a particular list and construct his own instrument on the basis of a principle derived from his examination.

In listing some forms here, we intended to attain the following goals with the reader:

- to make him familiar with widely used evaluation forms or checklists,
- to enable him to develop his own list of evaluation criteria to suit his own needs,
- to help him understand some basic principles underlying the construction of evaluation criteria lists.

Evaluating the Overall Programme of a School

The form presented in Exhibit 1 is taken from the criteria book of the National Study of Secondary School Evaluation (1960). It was designed for use by accreditation committees. The checklist is required to be completed by the members of the committee visiting a school on the basis of evidence provided by the school staff or gathered in the school during a site visit. The checklist in Exhibit 1 is an example of 80 different checklists used by the accreditation committee. While this form is designed to examine the overall organization of the school programme, the reader can adapt these items for evaluating any instructional material.

EXHIBIT 1: CHECKLIST FOR EVALUATING A PROGRAMME OF STUDIES

1. Organization

Checklist - The program of studies:

( ) 1. Is based upon the philosophy and objectives of the school.
( ) 2. Is based upon an analysis of the educational needs of youth.
( ) 3. Provides for both the common and individual needs of the students of the school.
( ) 4. Provides organized sequences of courses.
( ) 5. Provides courses or sequences of courses especially developed for the superior or advanced students.
( ) 6. Provides courses or sequences of courses especially developed for the slow-learning students.
( ) 7. Is flexible in time allotments in such ways as variation in number of periods for elective subjects, periods allotted to special help and remedial work, and time devoted to course work initiated by students.
( ) 8. Has prescriptive provisions to ensure a balanced program as to both scope and sequence for each student.
( ) 9. Recognizes the contributions made by the student activity program.
( ) 10. Is examined periodically to identify needed changes.
( ) 11.

Evaluations

( ) a. How well is the program of studies organized to meet the needs of different levels of ability?
( ) b. How well is the program of studies organized so that each student has a balanced educational program?
( ) c. How well does the program of studies provide for sequential study?

(National Study of Secondary School Evaluation, 1960)

General Evaluation Criteria

Eash developed a general form for evaluating instructional materials: various versions and editions of this form have been published and widely used (Walberg, 1974). The excerpt reproduced in Exhibit 2 is the ERIC Clearing-house variation of this instrument.

EXHIBIT 2: CONSTRUCTS OF CURRICULUM DESIGN

I. OBJECTIVES

A. What is the nature of the general goals of the material stated?

B. Are specific objectives stated for teacher use?

C. If neither of the above are stated, list what you believe are the intended objectives of the material.

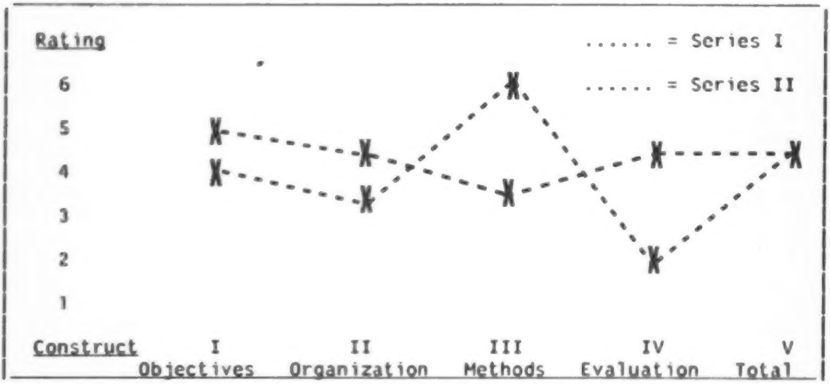
D. What are the main emphases in the objectives?

E. On the scale below, rate the objectives of the materials. Please circle an exact point.

Objectives not useful to a teacher.	1	2	3	4	5	6	Objectives give clear direction for instruction and are useful for a teacher.

In Eash's form four foci of the instructional materials are considered: objectives, organization of materials, methods of instruction, and instruments provided for evaluating learning. A completed form containing a comparative graphic representation of the evaluation outcomes of two sets of mathematics instructional materials is given in Exhibit 3.

EXHIBIT 3: A PROFILE OF TWO MATHEMATICS SERIES



The profiles in Exhibit 3 indicate that both mathematics series had equal overall rating but differed from each other with regard to all four unique foci of the instructional materials.

Guidelines for Evaluation

The guidelines reproduced in Exhibit 4 are extracts from a document developed by Armstrong (1976). The original document contains 25 guidelines.

EXHIBIT 4: EVALUATION GUIDELINES

Guideline 1. Materials should include complete and accurate identifying information.

This material includes complete and accurate information on the following:

- Title of Element (______________________________)
- Title of Programme/Series (______________________)
- Author(s) (__________________________________)
- Publisher/Producer address (_____________________)

Guideline 14. Materials should meet certain standards of technical quality.

This material has:

_____ Clear, legible, visible, visual presentations

_____ Clear, discriminable, sound presentations

Guideline 15. Materials should have up-to-date, accurate, and authoritative content and/or subject-matter.

A. This material has content and/or subject-matter which is:

_____ Accurate

_____ Factual

EXHIBIT 4: EVALUATION GUIDELINES (Continued)

____ Authoritative

____ Treats customs of different groups with dignity and respect

____ Provides personalized learning experiences

____ Demonstrates consideration for the human worth and dignity of all people

B. This material shows the diversity of value positions and/or philosophies. This material:

____ Presents a variety of value positions and philosophies

C. This material discourages stereotypic views and biases toward diverse human attributes, conditions, attitudes, and/or behaviors. This material:

____ Emphasizes the fact that every group has its achievers, thinkers, writers, artists, scientists, builders, and moral leaders

____ Avoids belittling anyone

The 25 guidelines contained in the document comprise hundreds of questions. While some guidelines (like No. 14 which deals with technical quality) require only a few marks, others (like No. 17, dealing with fairness and balance) contain 62 specific items. The 25 guidelines are clustered in five groups. The largest group describes the medium per se while the other clusters deal with media related to the curriculum, the media and their interaction with the learner, their interaction with the teacher and their interaction with the environment.

Criteria for Fairness

An increasing concern for eliminating bias and prejudice from instructional materials has generated criteria for fairness. Such criteria appear in a form developed by Rosenoff (1975). The items of the form are grouped in four major categories: validity, balance, unity and realism.

The basic set of criteria in Exhibit 5 is supposed to be supplemented by additional criteria which relate uniquely to the needs of the particular minority interest groups.

EXHIBIT 5: CRITERIA FOR FAIRNESS

Validity of Information

a. Information is accurate.

b. Stereotypes are not perpetuated.

Balance

a. A set of values in one culture is not described in such a way as to make them appear inferior in comparison to values in another culture.

b. The distinctive cultural characteristics of an ethnic group or groups are presented in a positive and comprehensive manner.

Unity

a. All points of view concerning historical events or issues are included (e.g. Is only the Anglo point of view emphasized? Other points of view should be emphasized as well).

b. The text presents the full range of events when discussing historical events.

Realism

a. Erroneous impressions are not created by citing selected facts and omitting others about an ethnic group.

b. Individuals and groups are portrayed in true-to-life fashion.

EXHIBIT 6: SIX CRITERIA CLUSTERS

No.	Desirability	Practicality	Intrinsic Quality	Product Development	User Effects	Spin-offs
1	Need	Adaptability	Sex, race, ethnic, age socio-economic fairness, or balance	Learner verification and revision	Cognitive skills, per-formance, attitude, motivation, learning rate, attendance, impact on others	Model for other work
2	Appropriate for intended users	Ease of use	Content accuracy and currency	Quality design	Credibility of evidence supporting claimed effects	Contributes to knowledge

EXHIBIT 6: SIX CRITERIA CLUSTERS (continued)

3	Adherence to social, moral and instructional values	Reasonable fiscal and psychic costs	Uniqueness	Expert staff		
4	Demand	Availability and acceptability to users	Instructional quality clear purpose, rationale			
5		Ability to receive any needed training to use product	Technical quality of physical features of materials			
6			Appeal to learners			

Clusters of Criteria

A concise set of 22 criteria grouped in six clusters (developed by Klein of the National Institute of Education, USA), in Exhibit 6 on the previous two pages, represents an attempt to produce an integrated approach to the evaluation of instructional materials (Woodbury, 1979).

The cluster titles are reminiscent of major evaluation types dealt with in professional literature such as Needs Assessment, Intrinsic Evaluation, Feasibility Study, Pay-off Study and Unintended Outcomes.

The Selection of Textbooks

An instrument to guide teachers in selecting social science textbooks was developed by Masako Tonaka (Woodbury, 1979). The whole instrument is reproduced in Exhibit 7.

EXHIBIT 7: ITEMS FROM A SELECTION GUIDE FOR TEXTBOOKS

Name of Textbook: ____________________ Grade: __________ Date: ________

Publisher: __________________________ Reviewer: ________________________

Author: _____________________________ School: __________________________

	Check one					
	Excellent	Good	Fair	Poor	No Opinion	Category* Scores
	3	2	1	0	0	
A. Racism and Sexism: Group 1 1. Illustrations (examples) a. stereotypes b. lifestyle c. tokenism						
2. Storyline a. relationships b. standard for success c. viewpoint d. sexism						
B. Authenticity 1. Accuracy of Facts/Appropriate to Context						

EXHIBIT 7: ITEMS FROM A SELECTION GUIDE FOR TEXTBOOKS (continue)

2. Impartiality of presentation						
3. Up-to-date Information						
C. Appropriateness 1. Contribution to the program objectives						
2. Vocabulary level						
D. Scope 1. Coverage of subject matter						
2. Concept development						
3. Skills development (map-reading, use of graphs, etc.)						
4. Process development (e.g., problem solving)						

EXHIBIT 7: ITEMS FROM A SELECTION GUIDE FOR TEXTBOOKS (continue)

E. Interest 1. Relationship to user's experience						
2. Intellectual challenge						
3. Appeal to students						
F. Organization 1. Sequential development of concepts						
2. Match with District sequences						

TOTAL SCORE = ______________________________

POSSIBLE RANGE = 0:123

Criteria Stated in the Form of Questions

Gall (1981) compiled a list of 39 evaluation criteria and grouped them into the following major categories: publication and cost, physical properties, content, instructional properties. He also formulated questions which could indicate whether a particular criterion had been met. The criteria together with the relevant questions are listed in Exhibit 8.

EXHIBIT 8: CRITERIA STATED AS QUESTIONS

Publication and Cost

1. Authors: Are the authors known and respected professionally?
2. Cost: Is the cost of the materials reasonable relative to other, comparable materials?
3. Development History: Were the materials adequately field tested and revised prior to publication?

Physical Properties

11. Aesthetic Appeal: Are the materials likely to appeal to the user's aesthetic sense?
12. Components: Do the materials contain so many components that teachers will have difficulty in keeping track of them and using them?
13. Consumables: Does the product make unnecessary use of consumable materials?
14. Durability: Do the materials have components that are specially vulnerable to wear?
15. Media: Does the developer make appropriate use of the media included among the materials?

EXHIBIT 8: CRITERIA STATED AS QUESTIONS (continue)

Content

18. Approach: Does the developer use an approach consistent with the district's curriculum?
19. Instructional Objectives: Are the materials' objectives compatible with the district's curriculum and acceptable to teachers?
20. Instructional Objectives - Types: Do the materials contain affective objectives in addition to cognitive objectives?

Instructional Properties

27. Comprehensibility: Will the materials be clearly understood by the students who will be
28. Co-ordination with the Curriculum: Are the materials compatible with other materials currently being used in the school?
29. Individualization: Does the design of the materials allow teachers to use them differently according to student needs?
32. Learner Characteristics: Are the materials appropriate for the students who will be using them?
33. Length: Are the materials an appropriate length so that they can fit conveniently into the teacher's instructional schedule?

A Form for the Analysis of History Textbooks

To determine trends in non-Western countries, a form for the analysis of textbooks in a large multilingual country is reproduced in Exhibit 9 (National Council of Educational Research and Training, 1972).

EXHIBIT 9: PROFORMA ANALYSIS FOR HISTORY TEXTBOOKS

ANALYSIS PROFORMA
(General Information)
History Textbook

1. Title of the textbook: ....
2. Grade level: ....
3. Language: ....
4. Author: ....
5. Publisher: ....
32. Total number of illustrations:
    (i) Photographs: ....
    (ii) Art work: ....
    (iii) Diagram: ....
    (iv) Graph: ....
    (v) Maps: ....
    (vi) Time line: ....
    (vii) Charts: ....
    (viii) Coloured: ....
    (ix) Black and White: ....
33. Variety of illustrations:
    (i) No. of decorative illustrations: ....

EXHIBIT 9: PROFORMA ANALYSIS FOR HISTORY TEXTBOOKS (continue)

(ii) No. of documentary illustrations: ....

(iii) No. of aesthetic illustrations: ....

39. Approach to the subject adopted in the book:

(i) Biographical (dynastic): ....

(ii) Sociological: ....

(iii) Both biographical and sociological: ......

(iv) Emphasis on the development of:

1. Political history: ....
2. Economic history: ....
3. Cultural history: ....
4. Administrative history: ....

40. Organizational pattern:

(i) Chronological sequence of events and happenings: ....

(ii) Thematic or Topical treatment: ....

(iii) Combination of (i) and (ii): ....

....

(iv) Nature of combination - (Emphasis on which aspect): ....

Criteria Combined with the Type of Data Needed

Below is a scheme for relating evaluation criteria to the type of data needed for providing evidence. Bloom (1977) suggested ten criteria for evaluating curriculum materials, and he also specified the type of data recommended for each of these criteria.

Bloom recommended judgemental, observational, and test data for appraising the worth of a particular set of curriculum materials.

EXHIBIT 10: CURRICULUM CRITERIA AND TYPES OF EVIDENCE NEEDED

Issues Being Considered	Judgmental Evidence				Observations in the Classroom	Student Learning
	Subject Experts	Other Specialists	Teachers and Supervisors	Students		
A. Accuracy and soundness of materials	X	X	X			
B. Internal structure and sequence	X	X	X			Criterion-referenced formative tests
C. Feasibility, cost and practicality	X	X	X		Problems posed for teacher training	
D. Clarity and meaningfulness	X	X	X	X		What has been learned well?

EXHIBIT 10: CURRICULUM CRITERIA AND TYPES OF EVIDENCE NEEDED (continue)

E. Level of difficulty	X	X	X	X	Student responses	Overall effectiveness
F. Interest value for students	X	X	X	X	Student attention and involvement	Interest and attitude questionnaires and scales
G. Special teaching problems	X	X	X		Conditions of instruction	What needs to be revised?
H. Relevance to students' needs and aspirations	X	X	X	X		
I. Overall reactions to materials	X	X	X			
J. Unintended learning and effects	X	X	X		General and specific comments by observers	

Criteria for Evaluating ETV-Programmes

As the major thrust of this book is on the evaluation of ETV-programmes, the criteria used for examining ETV-programmes will be dealt with in a separate section. Some examples of checklists will be presented and then criteria derived from research findings will be listed.

Criteria related to the formal features of TV will also be discussed.

FORMS FOR EVALUATING GENERAL EDUCATIONAL CHARACTERISTICS OF ETV-PROGRAMMES

Evaluation forms for ETV-programmes were published mostly in the years when such programmes were being widely disseminated. They follow a similar pattern to that for evaluating curriculum materials of other types, and show relatively limited consideration for the unique features of TV programmes. They focus on curricular aspects (such as objectives, appropriateness of demonstrations) and issues mostly dealt with in textbooks of psychology of learning (such as types of learning activity). An example of such an evaluation form was prepared by Roscoe C Brown Jr. It is presented in Exhibit 11 (Diamond, 1964).

The feedback questionnaires designed at South Carolina, Hagerstown Md and at Anaheim City School District also represent this overall curriculum-oriented approach (see Exhibits 12-14).

In the 1970's and 1980's criteria derived mostly from curriculum and learning theories, were used in evaluation research into the quality of ETV-programmes. Relatively little attention was given to the unique features of the medium.

EXHIBIT 11: ITEMS FROM A TV EVALUATION FORM

TV Evaluation Form

Lesson ______________________ Taught by ____________

Date ______________________ Grade ____________

Person rating ______________________

Please rate each question on the following scale:

1. Objectives

Lowest — Highest

a. Were the objectives clearly stated? 1 2 3 4 5

b. Were the objectives feasible for the level of students being taught? 1 2 3 4 5

2. The Lesson

a. How adequately was the lesson planned to achieve the objectives? 1 2 3 4 5

b. In your opinion how effective was the lesson planned for the level of students in your class? 1 2 3 4 5

Please rate the following aspects of the lesson:

	Lowest 1	2	3	4	Highest 5
Verbal presentation					
Use of visual aids					
Timing					
Appropriate use of demonstrations					
Effectiveness as a TV instructor					

EXHIBIT 12: ITEMS FROM A SOUTH CAROLINA TELEVISION EVALUATION FORM

SOUTH CAROLINA EDUCATIONAL TELEVISION CENTER
2712 Millwood Avenue
Columbia, S.C.

INFORMATION SHEET

Teacher ______________________________ Date ______________

School ____

Number of students in classroom ________ Subject ____________

TEACHING

Introduction:	Good	________	Comments:
	Fair	________	
	Weak	________	
Summary:	Effective	________	
	Ineffective	________	
Content:	Right amount	________	
	Too much	________	
	Too little	________	
Sequence:	Logical	________	
	Unrelated	________	
Vocabulary:	Good	________	
	Too difficult	________	
	Too easy	________	
Visuals:	Right amount	________	
	Too many	________	
	Too few	________	
Pacing:	Right	________	
	Too fast	________	
	Too slow	________	

Teaching technique: Strong points ____________________________

____

Weak points ____________________________

____

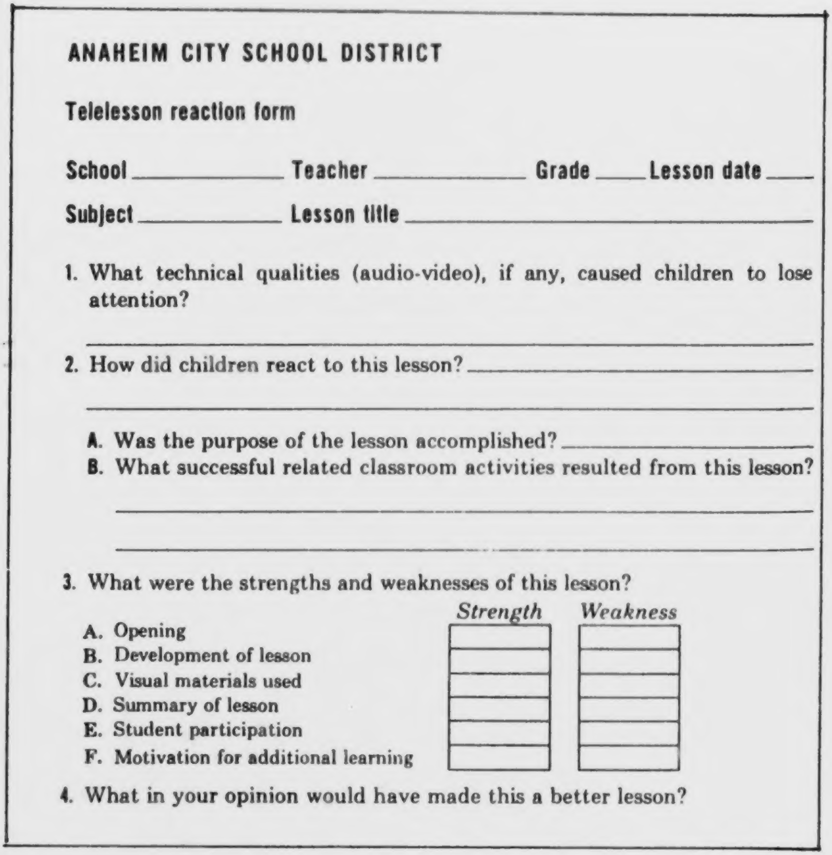
EXHIBIT 12: ITEMS FROM A SOUTH CAROLINA TELEVISION EVALUATION FORM
(continue)

What were the strenghts and weaknesses of this lesson?

	Strength	Weakness
A. Opening		
B. Development of lesson		
C. Visual materials used		
D. Summary of lesson		
E. Student participation		
F. Motivation for additional learning		

What in your opinion would have made this a better lesson?

EXHIBIT 13: ANAHEIM CITY EVALUATION FORM



ANAHEIM CITY SCHOOL DISTRICT

Telelesson reaction form

School ____________ **Teacher** ____________ **Grade** ____ **Lesson date** ____

Subject ____________ **Lesson title** ________________________

1. What technical qualities (audio-video), if any, caused children to lose attention?

____

2. How did children react to this lesson? ______________________

____

   A. Was the purpose of the lesson accomplished? ______________
   B. What successful related classroom activities resulted from this lesson?

   ____

   ____

3. What were the strengths and weaknesses of this lesson?

	Strength	*Weakness*
A. Opening	☐	☐
B. Development of lesson	☐	☐
C. Visual materials used	☐	☐
D. Summary of lesson	☐	☐
E. Student participation	☐	☐
F. Motivation for additional learning	☐	☐

4. What in your opinion would have made this a better lesson?

EXHIBIT 14: ITEMS FROM A FEEDBACK QUESTIONNAIRE

FEEDBACK QUESTIONNAIRE
Hagerstown, Md.

School________________________ Grade____________ Subject______________

Classroom teacher_______ Number of students_____ Lesson date___________

Level of student ability:
Above average __________ Average __________ Below average __________

1. Sound: Distinct___ Audible___ Not clear___ Describe difficulty___
2. Video: Clear___ Interference___ Describe difficulty___
3. Introduction: Stimulating___ All right___ Dull___
4. Development: Clear___ Interesting___ Weak___
5. Visuals: Enough___ Too many___ Too few___ Shown long enough___

   Not long enough___ (which one)___

   Effective___ Didn't make the point___ (which one)___
6. Speed of lesson: Right___ Too fast___ Too slow___
7. Vocabulary: Right___ Too difficult___ Too easy___

Salomon (1974) noted that the uniqueness of a particular medium is determined by the symbolic system it uses for encoding information, rather than by the technique used for transmitting the information. Seen from an educational point of view, a series of still pictures printed in a book differs from an ETV-programme transmitted through electronic devices: the mental functions carried out by the viewer are different. The differences relate to the types of materials provided to the learner. In addition, the differences in access to these materials are less significant from an educational point of view. Nevertheless, in the bulk of evaluation studies one seldom encounters media-unique criteria.

Lundgren (1972) lists a series of criteria for judging ETV-programmes. The general educational aspects of the teaching-learning process are predominant among them. Though he is aware that TV has unique possibilities, he merely mentions its uniqueness and describes it in very vague terms. One criterion listed by Lundgren is:

> *making use of the unique capabilities of the medium, i.e. trying to give the students subject matter (or content) that they would otherwise not get, and using forms of presentation they do not otherwise meet.*

In using this criterion Lundgren refers, for example, to travel programmes, where the different places are not easily accessible to the learners. Observing processes is another example: they can be more easily observed through the mediation of TV moving pictures than in reality. Examples of such processes may be the simultaneous operation of the valves of an engine, enlarged pictures of the operation of micro-organisms, a slow-motion presentation of human movements or observing an object from various angles, etc.

Among the criteria listed by Lundgren, one finds many which apply equally to all media used for teaching. Examples are:

- avoid instruction in a sense of telling facts
- provide experience for thought
- free creative energy
- select good presenters (a person we trust)
- use language the audience understands
- bring society to the classroom
- provide sufficient repetition (say something twice in different ways)
- link up with the experience of the audience
- do not impose ready-made opinions on viewers

Criteria Derived from Research Findings

Schramm (1972) summarizes research findings about ETV and provides a list of criteria which could be considered in examining ETV qualities. He groups them under four major categories: picture, sound, programme presentation and participation. Some examples of criteria in these four categories are listed below.

PICTURE - If a televised lecture is accompanied by pictures of a live audience, what aspects of audience behaviour should be presented to the TV viewer? What kind of background arrangements increase the effectiveness of an ETV lecture? No attempt will be made here to provide answers to these questions. They are mentioned here to indicate that research findings relevant to these issues may serve as a source of evaluation criteria for ETV-programmes.

SOUND - The speed of the speech (regular speech versus condensed or accelerated speech); background music; etc.

PROGRAMME PRESENTATION - Various aspects of the programme are examined. Should the programme be presented by classroom teachers, by trained actors, or by high profile, authoritative figures?

PARTICIPATION - The variables listed in this category are: covert participation of viewers (i.e. paying attention, carrying out certain mental operations) versus overt participation (doing something which can be externally observed, such as writing, drawing, speaking, moving or manipulating an instrument). Other participation variables are: feedback provided to the viewer; activities which serve as advance organizers for the programme versus practice and drill; spacing versus massing or blocking the practice; pausing; rest breaks and repetition.

Criteria Related to Formal Features of TV

Wright and Huston (1981) summarize findings related to the form, format, codes and conventions of television and their interaction with the child's growing cognitive and social competence. The authors strive to separate formal features of TV from content features of the TV programme with the aim of estimating the net effect of formal features of TV on the attainment of desired educational outcomes. The authors claim that formal features can be defined independently of content and that they can convey different kinds of messages, programme themes and plots. Three categories of formal features are identified: visual features such as cuts, pans, dissolves and special affects; auditory features such as music, sound effects dialogue; and modular characteristics such as pacing (rate of scene and character change), physical movement (action) and variation (p 66).

The form of children's programmes appears to be different from adult programmes: they have a higher rate of change and more salient images (intensity, contrast, novelty, incongruity, etc.).

Formal features of TV also help to distinguish scenes, bits, chapters, segments etc. within a single programme. They provide contextual information through long zooms, a moderate action level, etc. They establish assumptions about the off-camera environment and about the identity of people appearing on the screen. Features which capture attention are perceptually salient phenomena such as pace, movement, visual effects, music and non-human speech.

Formal features of TV lead viewers to certain mental processes, such as focusing on central incidents of the story, sequencing and integrating events. Formal features may also increase or decrease the amount of invested mental effort (AIME) needed to recognise visual content of an image on the screen, and to understand its meaning (Salomon, 1983).

An experienced TV viewer can infer the content of pictures without continuously looking at the screen, whereas a young child lacking such experience, may need to concentrate on the pictures to grasp their meaning. Wright and Huston (1981) report that salient features of the picture such as physical movement, and visual and auditory special effects can help identify central themes. Humour increases motivation to pay attention to a programme. Segmental marks facilitate comprehension, but not all artificial separators are easily understood by the child. Forms of signalling, such as a laugh track or jingle to announce a classification exercise are also useful. Advance organizers, brief previewing, or some kind of explanation may help the child understand novel features of a programme. Thus, Salomon (1979), for example, showed that using a long zoom to make the transition from a wide to a tight shot taught the child to understand the part-whole relationship implicit in a tight shot. Though the findings reported in this section are not stated in a form which directly defines evaluation criteria, they do illustrate the possibility of deriving evaluation criteria from findings of this type.

Attendance and Audience Reaction

A criterion frequently used in the evaluation of ETV and TV programmes in general is the amount of interest the target audience shows in a particular programme, and the actual viewing (attendance) of the programme. The extent of viewing can be measured in both the macro- and micro-context. Advanced technology enables us to estimate the number of viewers of a particular programme in a whole neighbourhood, and across the nation.

The estimation of the overall number of viewers as a criterion of the quality of a programme is more commonly used for summative evaluation, and for rating the popularity of various programmes. A detailed description of the technique used to measure the reactions of the audience and of problems related to using these techniques is given by Baggaley in this book. At the micro-level (for example, in a classroom or in a laboratory situation) one can determine the moment-by-moment level of attention to the programme, i.e. one can say which moment of the programme fascinated the audience and which left them indifferent or bored. The moment-by-moment measurement is commonly used for the purpose of formative evaluation when programme developers are in a position to change elements of the programme (Baggaley, 1985).

A variation of attendance or appeal studies is the observation of the viewer in a situation where competitive activities are offered. Examples of these are several TV programmes being presented simultaneously to the viewer, or toys, pictures and reading materials being left next to the TV set.

The Strength and Weakness of Visual Media

Deriving a particular piece of information from the verbal description of a situation, or alternatively from the visual representation of the same situation requires mental processes of different types (Olsen and Bruner, 1974).

Arnheim (1974) contrasted the unique features of deriving information from visual media with deriving the same information from verbal sources. He pointed out the advantages and limitations of visual media. Some of his observations, which may serve as a basis for evaluation criteria or ETV-programmes will be presented here.

The image of an object helps us to see things which we would otherwise probably not see. It helps to alleviate the indirectness of modern experience. The intuitive hunter had direct interaction with his prey. Contemporary man seldom deals directly with things which concern him. The presentation of images may raise reactions from the viewer, but it also may leave him a detached observer.

* The edited version of a recording, a learning experience, eliminates the uncertainties which create tension and involvement in an action. In transmitting an experiment in chemistry one misses the uncertainty of whether the experiment will succeed. "A cooking lesson transmitted live over television gains rather than loses by small mishaps ... The sense of adventure is a great asset in learning and the realisation that the teacher is not perfect provides legitimate encouragement" (Arnheim, 1974:186).

* An image should be presented in a way that answers a particular question or quiz. It is, therefore, good practice first to present a quiz or a question and give the mind the chance to answer, and then to supply the solution.

* Movement and change must be understood in relation to what remains unchanged.

* The time needed to explore a picture adequately varies ... It is necessary to give each scene sufficient time depending on its importance.

* The presentation of a series of pictures must be organized with sufficient lucidity. A child will not necessarily relate the different walls of the same building to a single object.

* The image one creates about the size of an object may be false. In studying art history, it is difficult to protect oneself from the distorting effects of a miniature painting appearing as large as Michelangelo's David.

* The part is different from the whole: if a picture conveys the message that leaves a child frightened as a reaction to his environment, then it is necessary to present the environment in the pictures as well.

* The eye does not record all the details of a picture. Familiarity with a situation may lead us to disregard the fact that a detail is missing in the picture. The adult seeing a picture of a Tyrolean playing the zither will understand the situation even if the strings are missing in the picture - the child will not.

* One has to integrate various details of a picture to understand the situation presented.

Concluding Remarks

This article has attempted to examine criteria used in evaluations across different disciplines. Our concern has been with evaluation as a discipline, done by evaluation experts. Evaluation activities done casually by non-experts in the course of their daily lives lies outside our field of interest.

Evaluation is a field of enquiry in many different disciplines. One finds evaluation studies in the social sciences, such as education, health, welfare, urban development.

Evaluation studies are a feature of the humanities, in art, literature, music; as well as of industry in quality control and consumer orientation. The legal sciences are also concerned with evaluations when they deal with the acceptability of evidence.

Evaluations are done for a variety of purposes using different procedures not only across disciplines, but within single disciplines. The music critic is engaged in the evaluation process whenever he writes a review of a recital. This evaluation, though it is a disciplined study, is quite different from an evaluation of an article written by a scholar in a scientific journal. Both activities are different from the task of a selection panel for a music competition.

Evaluation experts in different fields tend to develop their own canons de novo without paying attention to experience accumulated in evaluation activities in other areas. Nowadays, there are distinct theoretical and methodological developments in evaluation not only across broad areas of human concern such as education, literature, quality control in industry, but also within these areas. Educational evaluation which used to be a well structured discipline in the 1950's, has gradually become splintered into various sub-disciplines, each exhibiting a highly centrifugal tendency. Evaluation of student learning, of teaching in higher education institutions, of programmes, of educational environments, of educational media, have all moved speedily towards the status of autonomous disciplines. This development also characterises the evaluation of ETV-programmes. It is becoming an autonomous discipline with its own canons, its own group of distinguished experts, journals and conventions. While such a development may benefit the quality of ETV-programme evaluation, it may also lead to provincialism and narrow-mindedness. If ETV-programme evaluation is detached from the core of educational evaluation, it may mean that ETV evaluators will start 'to invent the wheel'.

In this paper the evaluation of ETV-programmes has been related to issues of programme evaluation in general, and an attempt has been made to demonstrate the relevance of evaluation procedures in general to the evaluation of ETV-programmes. The relevance of some evaluation theories, ideas and procedures outside the educational evaluation of ETV-programmes was also examined.

The production of an ETV-programme requires the close co-operation of professionals and artists of different types. The evaluator of the programme should consider various aspects of ETV-programmes as they are viewed by persons with different kinds of expertise. For example, subject-matter specialists, educational psychologists, curriculum experts, literary critics, communication specialists, television critics, critics of acting, film critics, are some of the professionals whose views should be sought. Their views could be integrated and used to develop criteria for evaluating ETV-programmes.

This article has dealt with only some of the views of the ETV-programme specialists. The full integration of all relevant evaluation criteria for ETV-programme evaluation into a single scheme would require a well-planned collaborative study of specialists in the different fields.

REFERENCES

Adar, L and Fox, S 1978. Analysis of a History Curriculum and of its Implementation in Schools (Hebrew). School of Education, The Hebrew University, Jerusalem.

American Educational Research Association and National Council on Measurements used in Education 1955. Technical Recommendations for Achievement Tests, National Education Association, Washington D C.

American Psychological Association 1966. Standards for Educational and Psychological Tests and Manuals, American Psychological Association, Washington D C.

Armstrong, J R 1976. Guidelines used in consumer information analysis. Paper presented at the American Educational Research Association Annual Conference.

Arnheim, R 1974. Virtues and vices of visual media. D R Olsen (ed.), Media and Symbols: The Forms of Expression, Communication and Education. National Society for Studies of Education Yearbook vs 73rd part I, Chicago, University of Chicago Press, 180-210.

Baggaley, J 1985. Formative evaluation of educational television. Educational Programmes RTV1, Vol. 1(1), 48:63.

Ben Peretz, M 1985. Curriculum potential. T Husen and T N Postlethwaite (ed.), International Encyclopedia of Education: Research and Studies. Oxford, Pergamon Press.

Bloom, B S 1977. Curriculum try-out. A Lewy (ed.), Handbook of Curriculum Evaluation. New York, Longman.

Braskamp, L, Brandenbury, D C and Ory, T C 1984. Evaluating Teaching Effectiveness. Beverley Hills, Sage.

Cambre, M A 1981. Historical overview of formative evaluation of instructional media products. Educational, communication and technology - A Journal of Theory, Research and Development, 29(1):3-25.

Choppin, B 1978. Item Banking and the Monitoring of Achievement. National Foundation for Educational Research, Slough, England.

Devereux, F L 1933. The Educational Talking Picture. Chicago, University of Chicago Press.

Diamond, R M 1964. A Guide to Instructional Television. New York, McGraw Hill.

Eden, S 1984. Evaluation of Learning Materials (in Hebrew, English Summary). Curriculum Centre, Ministry of Education and Culture, Jerusalem.

Eraut, M, Len G and Smith, G 1975. The Analysis of Curriculum Materials. Department of Education, University of Sussex, England.

Eraut, M 1985. Intrinsic Evaluation. T Husen and T N Postlethwaite (eds.), International Encyclopedia of Education. Oxford, Pergamon Press.

Gall, M D 1981. Handbook for Evaluating and Selecting Curriculum Materials. Boston, Allyn and Bacon.

Glass, G V 1969. The Growth of Evaluation Methodology. Laboratory of Educational Research, University of Colorado, Boulder.

Glass, G V 1978. Standards and Criteria. Journal of Educational Measurement, Vol. 15, 237:61.

Harrison, B 1968. Some uses of "Good" in criticism. F J Coleman (ed.), Aesthetics. New York, McGraw Hill.

Henze, D F 1968. The work of art. F J Coleman (ed.), Aesthetics. New York, McGraw Hill.

Hirsch, E D 1976. The Aims of Interpretation. Chicago, University of Chicago Press.

Hirsch, E D Jr 1969. Privileged criteria in literature. J Sterelka (ed.), Problems in Literary Criticism. Pennsylvania State University, University Park.

Hospers, J (ed.) 1969. Introductory Readings in Aesthetics. The Free Press, New York.

Joint Committee on Standards for Educational Evaluation 1981. Standards for Evaluation of Educational Programmes. New York, McGraw Hill.

Kaiser, A 1983. Legitimationmodelle in der Curriculum Entwicklung. V Hameyer et al (ed.), Handbuch der Curriculumforschung. Beltz, Weinheim.

Kenny, A 1967. Criterion. P Edwards (ed.), The Encyclopedia of Philosophy. New York, MacMillan.

Lungren, R 1972. What is a good instructional programme? W Schramm (ed.), Quality of Instructional Television. Honolulu, University of Hawaii Press.

Mager, R F 1962. Preparing Instructional Objectives. Paul Alto, Feardon Publishers.

Mann, J S 1978. Curriculum criticism. Curriculum Theory Network, Issue 2, Winter 1969/9, p 2-14. Reprinted Willis, G (ed.), Qualitative Evaluation. Berkely, McCutchen.

Medley, D E 1985. Evaluating teaching: criteria. T Husen and T N Poslethwaite (eds.), International Editors. International Encyclopedia of Education: Research and Studies. Oxford, Pergamon Press.

National Council of Educational Research and Training 1972. Preparation and Evaluation of Textbooks in History. New Delhi, Ncert.

National study of Secondary School Evaluation 1960. Evaluation Criteria. National Study of Secondary School Evaluation, Washington.

Olsen, D R and Bruner, J F 1974. Learning through experience and learning through media. Olsen D R (ed.), Media and Symbols: The forms of Expression, Communication and Education. National Society for Study and Education, 73rd Yearbook, University of Chicago Press, Chicago, 125-150.

Rieken, H W and Boruch, R F 1974. Experimentation: A Method of Planning and Evaluating Social Intervention. New York, Academic Press.

Rossenhoff, W E 1975. Sample Criteria for Fairness San Francisco, Far West Regional Laboratory for Educational Research and Development, October. ED 114-213.

Rossi, P and Freeman, H E 1982. Evaluation: A Systematic Approach. Berkely, Sage.

Schramm, W (ed.) 1972. Quality of Instructional Television. Honolulu, University of Hawaii Press.

Salomon, G 1974. What is learned and what is taught: the interaction between media, task and learner. Olsen D R (ed.), Media and Symbols: The Forms of Expression, Communication and Education, National Society for Study of Education, 73rd Yearbook I. Chicago, University of Chicago Press.

Salomon, G 1983. Television watching and mental effort. J Bryant and D R Anderson (eds.), Children's Understanding of Television. New York, Academic Press.

Salomon, G 1979. Interaction of Media, Cognition and Learning. San Francisco, Jossey Bass.

Scriven, M 1972. Prose and cons of goal-free evaluation. Evaluation Comment, Vol. 2, 28-33.

Shepherd, R S 1972. A checklist for analysing the style of instructional materials. Research 1972, Vol. 2, 28-33.

Stake, R 1967. Countenance of educational evaluation. Teachers College Record, Vol. 68, 523:40.

Tyler, R W 1949. Basic Principles of Curriculum and Instruction. Chicago, University of Chicago Press.

Tyler, R W 1973. Teaching for accountability. A C Ornstein (ed.), Accountability for Teachers and School Administrators. Belmont, Feardon Publishers.

Walberg, H J (ed.) 1974. Evaluating Educational Performance. Berkely, McCutchen.

Wheeler, C W 1980. National Council for the Accreditation of Teacher Education: Does it matter? East Lansing, MI, Institute of Research in Teaching.

Willis, G (ed.) 1978. Qualitative Evaluation. Berkely, McCutchen.

Woodbury, M 1979. Selecting Materials for Instruction: Issues and Policies. Libraries Unlimited, Littletown Company.

Wright, J C and Huston, A C 1981. Children's understanding of the forms of television. M H Kelly and H Gardner (eds.), Viewing Children through Television. San Francisco, Jossey Bass.

Part Two: METHOD

COMPLEXITIES OF SUMMATIVE ETV EVALUATION: POSSIBLE SOLUTIONS THROUGH A DIAGNOSTIC APPROACH

G Salomon

So, we have educational television (ETV) with a variety of programmes and accompanying materials to be broadcast to schools and homes, and we want to evaluate the merits and worth of a programme. Quite a noble and well-intended idea. And as evaluation projects are in abundance providing possibly useful models (e.g., Ball and Bogatz, 1971; Mayo, Hornik, and McAnany, 1976; Smith and Bissel, 1970), we should not face much difficulty in specifying the objectives, criteria and methods most suitable for such a project. Things, however, as we shall see later on, are far more complex (e.g., Shavelson and Webb, 1986).

Intuitively and niaively thinking, what is it that we would want to learn from an evaluation of an ETV-programme? In essence, what the evaluation of an ETV-programme or set of programmes would aim at, would be to answer such questions as to whether the programme achieves its goals, whether it has an effect, and what its relative effectiveness is, given some viable alternative (Cronbach and Associates, 1980). Toward these ends we would design a plan, which, depending on specific local possibilities, would use some experimental components; and carry out the plan. For example, if possible, we would tend to broadcast the programme(s) to some schools or homes, but not to others, thus obtaining some semblance of an experimental design. If impossible, we would perhaps use a time-series design (Campbell, 1969), age cohorts (Ball and Bogatz, 1971) or other quasi-experimental methods.

This description however, while applicable to the commonly employed evaluations of ETV and other educational undertakings, hides great complexities and difficulties. Some of these complexities challenge the basic assumptions underlying such apparently simple modes of ETV evaluation, often rendering them useless or even meaningless (Cronbach and Associates, 1980). In what follows, the specific complexities of ETV evaluation are presented. In their light, it should become evident why one cannot escape raising far more complex evaluation questions about ETV, questions that pertain to such issues as to what programme, instruction and implementation elements affect what kind of learner, under what kind of situation, and the like. Following this, the initial idea of evaluating ETV will be revisited and the concept of diagnostic evaluation will be introduced. Finally, an evaluation plan that attempts to deal with the described complexities by combining different kinds of evaluations into one plan within the general framework of diagnostic evaluation, will be presented to serve as an illustration.

Specific Complexities and Challenges of ETV Evaluation

The complexities to be seen in summative ETV evaluation stem, in large part, from the very nature of educational television. They are not always unique to ETV, but play a prominent role in its evaluation. Let us examine the major complexities in some detail.

The 'it' in 'Does it make a difference?'

The basic question one usually wants to address is 'What difference does a particular programme (or a number of programmes, or ETV in general) make in the light of its (or their) goals?'

One (niaively) assumes that the independent variable, the 'treatment', the 'input', (say, a programme) is specifiable as a clear and well defined entity, the impact of which can be measured. But what is a 'programme?' Is it the (often new) curriculum delivered via ETV that is supposed to 'make a difference?' Is it the accompanying materials? The special teacher training courses provided? The excitement of a new medium now introduced into the classroom, or home? Or is it the television broadcast itself that is supposed to carry the main responsibility for change? The answer to such questions, one might argue, is relatively simple: ETV presents educational 'treatments' that entail numerous, interrelated components, and these should be treated as whole packages. But, as we shall see later on, the evaluation of 'whole packages' is extremely complex, for the boundaries are neither clear nor is the entity - programme - well defined. (Has the didactic intergration of the ETV materials into the regular teaching part of the 'package' been effected?) Yet, as Leinhardt (1980) and others have pointed out, "in an evaluation, it is important to attribute educational gain to an aspect of an education treatment" (p 396).

Moreover, the evaluation of complex educational packages entails a paradox of important consequences, a point raised by Papert (1985) with respect to the evaluation of computer-related learning activities. An ETV-programme is not an isolated entity to be compared with some alternative where 'all other things are held constant'. If it is to have an effect of some value then its introduction has to be accompanied by a host of additional factors - for example, different modes of teaching the material in class, or different ways of televiewing at home. These are necessary, even crucial parts of 'the package' itself. But they are not, they cannot be, uniform across schools, classes, teachers, or homes. Thus, as Shavelson and Webb (1986) point out, "... the treatment cannot be defined independently of the constellation of characteristics in the local implementation, and multiple telecourse treatments (thus) abound" (p 4).

On the one hand, the more such factors involved - the better the chance that the programme might have the desired effects. But how do you evaluate the programme together with such a diversity of factors when nothing can be said to be kept constant? So, you want to evaluate the 'core' of the package, leaving out all those additional factors that, although important, are - from an evaluation point of view - just 'noise'. If nothing else, at least that 'core' is relatively uniform (so one thinks). But then the evaluation of 'the package', stripped down to its most basic elements (the ETV-programme, perhaps also its accompanying materials) may reveal very little; it leaves out some of the most crucial elements. This poses a serious dilemma for summative evaluation, for it undermines the assumption that you have a clear and definable 'treatment'.

What is Broadcast and What is Received?

A second complexity pertains to the many ways audiences are exposed to ETV-programmes and to the many ways in which they either choose to, or are led to, deal with them. One often (and implicitly) assumes that the programme, once produced and sent out over the air, constitutes a more or less uniform experience. But having a programme on the air and experiencing it are two different events. First, there is great diversity in the sheer amount and consistency of in-schools or at-home exposure to a given programme. Some teachers make systematic use of a programme, others expose their students to it only occasionally or rarely (e.g., Stasz and Schlossman, 1986). Second, there is an equally great diversity in the qualitative ways in which a programme is utilized by teachers, parents or the learners themselves. Whereas some teachers integrate ETV-programmes into their regular teaching, others use the programmes as peripheral add-ons. Similarly, whereas in some schools or homes children expend much mental effort in a programme, mindfully relating its contents to already acquired knowledge or to real-life events, others perceive the programmes to be mentally undemanding, just a fun activity, thus dealing with them with relative mindlessness (Salomon, 1983: Salomon and Leigh, 1984).

It becomes apparent that the ways in which educational television is used are of paramount importance. Such diversities are not random. The teachers that make more systematic use of ETV, the homes in which ETV is taken more seriously, and the learners who expend more mental effort in processing the presented information, are often the ones that may need these programmes the least, or are the best prepared for them. Hence they derive the most from these programmes (e.g., Campbell, 1969; Cook, et al, 1975). It becomes apparent that no evaluation of ETV can be limited to the programmes themselves; processes of implementation and utilization, their quantities and qualities are as much part of the package as are the materials themselves (see Leinhardt, 1980 for a detailed discussion of implementation evaluation).

ETV like a Drop in the Ocean

When a new curriculum in, say, biology is introduced and properly implemented one can assume with relative confidence that the observed improvements must be a direct or indirect result of the implementation. The curriculum, you might say, provides more or less most of what the students encounter in this domain. Not so with most ETV-programmes. ETV, in general, is an addition to ongoing educational processes, whether in schools or in homes. ETV does not monopolise, so to speak, any domain of knowledge. As already observed by Schramm (1977; Schramm et al, 1967), the better developed the educational system, the less of an observable impact does ETV have on it. Indeed, the learning outcomes that Ball and Bogatz (1971) attributed to Sesame Street in the U.S., where most children at that time did not attent pre-schools, were far larger than those observed in Israel (Salomon, 1976) where the majority of children did attend pre-schools. On the other hand, ETV was found to have a noticeable effect in a country such as El-Salvador, but then it was the whole educational reform that was introduced that had the observed impact rather than ETV itself (Mayo, et al, 1973).

Knowing that, one asks the question as to the size of learning effects that could reasonably be attributed to ETV-programmes, in and of themselves, when used in an otherwise well-developed educational system. It may not be more than a drop in the ocean. One faces here a great risk of committing an error of the second type whereby no overall effect is observed simply because the 'treatment' is too weak relative to the context in which it is implemented. This, it appears, is the conclusion to be derived from the evaluation of Think About, a programme designed to foster thinking skills in fifth and sixth graders in the U.S. Ostensibly, its overall effects were very weak (Sanders, 1983), illustrating the drop in the ocean phenomenon. However, few interactions where stronger effects could be found, were tested.

Common and Unique Effects

The 'drop in the ocean' phenomenon may conceal another, more fundamental complexity of ETV evaluation. What are the outcomes to be evaluated? Frequently programme goals and objectives are formulated in terms of knowledge acquisition (or in rarer cases - attitude change) that could be compared with achievements attained by other means. ETV is assumed to be just another means to the same educational ends (Olson, 1974): better knowledge of the sciences, better mastery of a new language, better acquaintance with career choices, better understanding of world affairs, and the like. But do such objectives exhaust ETV's possibly unique effects, the ones that are not common to other, non-television means?

Television, indeed, is more than just a convenient alternative delivery system. The medium offers the choice of symbolic modes of communication that, if properly used, cannot be duplicated by other means. The medium can present information in ways that address learners' cognitions in rather unique ways (Salomon, 1979), thus possibly attaining unique learning outcomes.

Consider, for example, the possibility that television's pictoriality creates 'bushier' associative networks than, say, language or print (Bagget, 1984), that it can arouse emotions that no other medium would (e.g., Gombrich, 1974), or that it can cultivate figurative internal modes of cognitive representation when mental operations are overtly supplanted on the screen (Salomon, 1979). These are not some unintended 'side effects' that evaluation should be generous enough to include; these are possibly the main effects to be measured, respecting the possibility that ETV and its alternatives are not just alternative routes to the attainment of the same ends.

There are many ways of understanding Newtonian motion or career choices, and there are numerous ways to represent cognitively such knowledge to oneself (e.g., Kosslyn, 1981). One mode need not be 'better' than another on average, just different and more appropriate for some learners. One way may be bettersuited for some kinds of learners, for example - 'visualisers', while another alternative - for other learners, perhaps the 'verbalisers'. On average the attainments could appear to be similar, but in actuality they are not.

Imagine an ETV-programme that depicts in visual, partly contrived and animated form, the way the human brain functions. Exposure to such a programme may lead, on average, to the same knowledge gains as other instructional means when a common measure of achievement is employed. But there is surely more to learning from the visual presentation than that. Proper evaluation of the programme would need to examine that non-common aspect, perhaps by means of ingeniously devised measures, to discover the unique impact, of the programme (e.g., how the information is represented in cognition or how resistant to change it is). The more specifically tailored such measures are to the nature of ETV - the less does the 'drop in the ocean' phenomenon obscure results. But then, such unique outcomes, because they are unique, allow no comparison with alternatives, hence one cannot say whether this or that ETV-programme is 'better' than its alternatives. One can only say what that programme is capable of achieving with one kind of learner or another.

The Diversity of Outcomes

The question pertaining to the nature of the outcomes to be evaluated has still another aspect to it. Learning, whether common to other means of instruction or ETV-specific, is one set of outcomes; it is usually the set most evaluations are concerned with. But there are other possible effects or outcomes worth examining, effects the importance of which may far exceed straight learning outcomes. Consider a few hypothetical examples. As a result of introducing a new ETV-programme in science, a whole new wave of interest in the domain develops in the school; more students want to learn science and more begin to consider careers in the area. In another place, teachers begin to think more seriously about the way they teach, feeling they have been drawn into a competition with ETV, a competition they would not want to lose (or, in other cases - a competition that exacerbates their perceived poor self efficacy). Or consider the possibility that with the introduction of ETV-programmes for pre-schoolers, some families for whom television no more than an electronic babysitter, start considering it as an educational tool deserving serious attention. Indeed, one of the more interesting results of broadcasting ThinkAbout was its effects on the thinking-skill curricula in the schools (Sanders, 1983).

Such possibilities are often not part of the initial list of programme objectives and are therefore not considered targets of evaluation. In other cases they are perceived as 'side effects' and somehow added to what is considered the main targets of the evaluation. The point here is that given an undefined treatment which is implemented and handled in a variety of ways, a variety of outcomes can also be expected. It is not just a matter of anticipating the unexpected (Cook et al, 1975); it is a matter of conducting ETV evaluation more as a widely spanned exploration than as a product inspection (Cronbach, 1982).

The Evaluation of ETV Revisited

Given such complexities, why should summative ETV evaluation be carried out? What purposes and functions is it to serve? What should it be a summative evaluation of? How should it be designed? At first sight, this may seem a rather odd set of questions. ETV needs to be evaluated because one wants to know if it has the desired and planned effects, and what its relative effectiveness is. Alternatively, one may want to consider the possibility of improving a programme, the accompanying materials, the delivery procedures, the implementation, etc., or even re-doing them. This, however, requires a far better understanding of what in the programme or its implementation affects whom, in what way and under what conditions something that simple summative evaluation cannot provide.

Indeed, if, as the preceding section suggests, the nature of the evaluation object is unclear and highly diversified in its usages; if numerous accompanying factors, more than the evaluation object itself, may account for learning outcomes; if there is a great quantitative and qualitative variability in the ways ETV products are used and experienced; if ETV, in and of itself, may often be no more than a drop in the ocean; if its main impacts may be unique but subtle; if outcomes are numerous and highly diversified; in short, if numerous known and unknown interacting factors are part of the process; then no simple model of summative evaluation applies. ***IT BECOMES EVIDENT THAT SIMPLE SUMMATIVE EVALUATION WILL NOT SERVE US WELL BECAUSE IT DOES NOT TAKE INTO CONSIDERATION THE KINDS OF COMPLEXITIES AND DIVERSITIES DISCUSSED ABOVE.*** In fact, hardly any of the implicit assumptions underlying simple outcome evaluation appear to apply. The idea with which we have begun would need to be revisited.

Barring questions motivated by politics and public-relations, what is it that one really would want to learn from ETV evaluation? What would the real purpose of the evaluation be when the complexities discussed above are seriously taken into consideration? Surely not to pass general judgement on a programme based on averages of improved achievement scores. As Cronbach (1982) has pointed out -

> *Most educational programmes are loosely defined, allowing room for adaptation at the state, community, and classroom level. An estimate of average or global benefit is only moderately helpful to national planners..., and is much less helpful to local decision makers. No group concerned with a particular site wants to know how a program has operated in some mythical 'average' site. 'How will it work for us?' Is the question (p 14).*

More likely, one would want to gather data to answer other than judgemental questions such as (a) *HOW IS A PROGRAMME USED, BY WHOM, IN WHAT WAYS AND WITH WHAT EFFECTS*, and (b) *WHAT COMBINATION OF FACTORS, INCLUDING THOSE THAT CHARACTERISE THE PROGRAMME ITSELF, ITS DIVERSIFIED MODES OF IMPLEMENTATION AND UTILIZATION, THE LEARNERS' MAKEUP, THE SOCIAL AND ORGANIZATIONAL SETTINGS. AND THE LIKE, PROMOTES, OR HINDERS POSITIVE (AND NEGATIVE) EFFECTS*. In other words, one would want to answer DIAGNOSTIC questions of the following general kind: ,

> *To what extent, by what means and in what ways, does a particular ETV-programme with such and such characteristics have what kinds of effects with what kind of learner and setting?*

Diagnostic Evaluation

I call this kind of approach diagnostic evaluation because it is not designed to yield a general verdict on a programme, a verdict that is unlikely to affect decision making anyway (Cronbach et al, 1980). Nor is diagnostic evaluation designed to provide basic, generalizable and scientific conclusions (although these would of course be welcome). It is also not designed to study whether one particular factor (e.g., classroom discussions following viewing; see Johnston and Ettema, 1982) increases the effects of a programme.

Rather, *ITS PURPOSE IS TO SHOW US QUITE PRECISELY, THROUGH A SERIES OF PLANNED STEPS OF INQUIRY, WHERE, HOW, WHY AND WITH WHOM PARTICULAR ELEMENTS OF THE ENTIRE PACKAGE HAVE POSITIVE OR NEGATIVE EFFECTS.*

Diagnostic evaluation, as conceived here, is based to some extent on ideas developed in the context of formative evaluation (e.g., Palmer, 1974; Scriven, 1973) and in the context of implementation evaluation (e.g., Leinhardt, 1980). Formative evaluation attempts to identify programme elements and exposure situations that need improvement, yielding information that can be fed back to designers and producers long before the whole package is fully implemented. Diagnostic evaluation has in principle the same objective in mind, but differs from formative evaluation in two important ways: it is carried out as part of summative evaluation, thus serving as a retrospective diagnosis, and it is an evaluation of the whole ETV package as implemented and utilized in 'the real world', rather than under semi-controlled and isolated conditions.

Implementation evaluation attempts to relate amounts and modes of implementation of a programme to learning outcomes, something shared also by diagnostic evaluation. But implementation evaluations "will not clarify why one set of approaches work(s) better than another, nor are they likely to advance our understanding of the relationships between compensating instructional processes" (Leinhardt, 1980:416), something that diagnostic evaluation attempts to do. A number of assumptions underlie the idea of diagnostic evaluation. They are enumerated below only briefly as they derive from the complexities discussed above.

* *If a product is to be improved (or if a new, improved product is to be designed), one would need to know with some moderate confidence what exactly needs to be improved and why. The observation that a TV maths programme, part of a distance learning course for adults, has little effect on achievements may result from its poor (or too abstract) quality, from the fact that it is broadcast at 06h00, or because it is unrelated to the textual material.*

Only when the evaluation is carried out in the actual ecological context can this identification take place.

- *No ETV product is all good or all bad for everybody or for everything. Certain elements may have positive effects, other elements may be neutral and still others may be harmful for some, but not for other learners, or in some, but not other settings. The assumption is that the effectiveness of as complex a treatment as that provided by ETV results from combinations of interacting factors such as population, programme, didactics, implementation, and setting: "the quest for an effect 'free and clear' of other effects (i.e., so called main effects) is unrealistic" (Cronbach, 1982:32). The purpose of diagnostic evaluation is to point these combinations out.*

- *For diagnostic evaluation diversities of components, settings, modes of utilization, modalities of implementation, learners, backgrounds, possible outcomes, and the like are an asset rather than a liability.*

- *Some factors that may account for ETV outcomes are more proximal than others and one would need to consider both proximal and distal factors; as is often the case, once the latter are examined (SES in the case of the effects of television on reading ability; e.g. Hornik, 1981) the contribution of the proximal effects may shrink significantly (Scarr, 1985).*

- *The various relevant factors come as varying "clouds of correlated events" (Scarr, 1985), and they can also shift roles: what is an independent variable in one case (e.g. preparing students for a broadcast) can be seen as a mediating one in another; and what is a mediating variable in one setting (parents' involvement in explaining a programme for their pre-schooler), may be taken as a dependent variable in another. Variables affect each other reciprocally (Bandura, 1982). Thus, the intrinsic qualities of a programme may affect the depth with which students process its information, but the depth of processing may have a reciprocal effect on the programme qualities that students notice and attend to. ETV evaluation needs to take such possible reciprocities into consideration.*

- *Proper diagnostic evaluation cannot be carried out in a single study; it is a matter of stepwise inquiries going from a general and descriptive study to the study of the specific that allows causal inferences.*

An Illustration

An illustration of a stepwise diagnostic evaluation plan for ETV follows. It was designed for the evaluation of the Israeli Center for Educational Television but was not actually carried out for personal and political reasons. It was during the planning of this evaluation that the concept of diagnostic evaluation was conceived and refined. Although the plan pertains to the evaluation of ETV in general, it is applicable, given a few obvious changes, to the evaluation of single programmes or series of programmes as well.

As with diagnoses in general, one wants to narrow one's focus gradually, basing each step of the process on the preceding steps. One also wants to examine key elements separately, lest they get lost among a host of other factors. Thus, the various stages of diagnostic evaluation follow a cone-like (or funnel) progression.

For purposes of temporary convenience we might divide the variables relevant to evaluation into four gross groups:

- *Relevant background variables that include both distal variables such as the kind of educational system and setting (large, small, urban, rural), as well as more proximal ones such as teachers (or home) and learners' characteristics, prior mastery of the knowledge to be learned, initial attitudes towards the programme, purposes of exposure to the programme, and the like.*

- *Input variables that include the programme(s) to be evaluated; their specific characteristics (e.g., extent of humour, didactiveness, degree of 'televisionness', degree of intergration between teaching and entertainment, dependency of one broadcast unit on preceding units, etc.); the nature of the accompanying materials and their relationship to the broadcasts; congruity between declared goals and their realisation in the programme, and the like.*

- *Process variables that include amount of in-class or at-home usage; degree of systematic usage; mode of the programme utilization (is it integrated or added to on-going activities, student preparation for a broadcast, etc.); attention; the mental effort expended in processing the programme; activities that follow the exposure, and the like.*

Outcome variables that pertain to learners (learning achievements, attitudes, abilities, self-perceptions, aspirations, beliefs, etc.); to teachers or parents, to schools (e.g., changed curricula), and to school systems.

This division is just a well accepted matter of convenience, but as stated above, any one of these variables or groups of variables could easily be positioned elsewhere, depending on the specific purposes of the diagnostic evaluation in the series of evaluation steps.

The First Diagnostic Step: Mapping the Terrain

With this tentative list of variables in mind we could now lay out the questions of the first and most general step in the series of evaluation studies. These questions are best represented in the form of a 'mapping sentence' as follows:

The amount of variance accounted for by:

A
- Amount of use
- Consistency
- Kind of use
- Length of use

of: ....

B
- ETV programme
- Written material
- Teacher training
- Promotion

designed for:

C
- School
- Home
- Both

in the area:

D
- Science
- Language
- Careers
- Thinking

for the age group:

E
- 3-5
- 6-9
- .
- .
- Adult

with attributes:

F
- Humour
- Didactive
- Entertain.
- Integration

in: ....

G
- Large urban centres
- Small townships
- Rural villages
- Suburbia

on measures of:

H
- Specific achiev.
- General achiev.
- Attitudes
- Abilities
- Behaviours
- Self-percepts.

of:

I
- student
- teacher
- parent

who are high/low on initial measures of:

J
- General ability
- Specific knowledge
- Specific abilities
- Learning attitudes
- Home background

As it can be seen, there are ten facets, but their number could, of course, be larger or smaller, depending on the kinds of elements, or facets that one would want to include or exclude.

Obviously, when only a single programme is to be evaluated various facets of this mapping sentence become irrelevant as they entail no variability. It will be noted that such a mapping sentence allows for the creation of various combinations of components across facets. And each combination would be compared with other combinations, making maximal use of the 'naturally given' variability. Clearly, not all combinations apply, and not all combinations need to be studied. But, given enough variation, one would want to focus on those combinations of greatest interest.

Consider the evaluation of a course in ecology to be used in school, designed for ninth graders. The following would be a reasonable list of combinations of variables, the contribution of which to a variety of outcomes would be evaluated through a series of comparisons: how much variance on measures of specific comprehension of ecology, subjective confidence in the knowledge, and attitudes towards ecology (facet H) is accounted for with students high vs. low in SES, and initial mastery of ecology concepts (facet J) when provided with smaller vs. greater amounts of programme use (facet A) with more vs. less teacher training (facet B), when the programmes are more visually saturated vs. more verbal (facet F), when used in large urban vs. small rural schools (facet G)? (Comparisons of extremes are presented here for illustrative purposes only; usually one would use more graded continua and correlational methods such as multiple regressions).

This phase of the evaluation can, of course, be the only one, with no subsequent phases following it. It is based on a wide sampling of classes and schools, and taps variability on a number of important variables, thus it can yield most of what summative evaluation, with a touch of diagnostics, would aim at.

If properly carried out, one would be able to determine whether an amount of exposure, on its own or in interaction with, say, written materials, teacher training, school settings and relevant individual differences, accounts for desirable results in one or more of the measured criteria with students at various levels of prior mastery.

Carrying out this phase of evaluation is relatively easy as data are collected by means of questionnaires and test materials sent out to the schools. But this phase yields no data on other important variables the measurement of which is costly (e.g., how is the programme used), but may not require the large number of classes (or homes) needed for the present phase. For example, one would want to know why some schools (or homes) make more or make less use of the programme. One would also want to know how much qualitatively different modes of using the programme, perhaps in interaction with other variables, contribute to the designated outcomes. This phase of the evaluation cannot yield data on such issues.

The Second Diagnostic Phase: Reasons for Heavier Use and its Effects

The first phase of the evaluation has shown us who are the heavier and who are the lighter users of the programme. In the second phase we narrow the scope of the evaluation and the size of the studied sample to examine what accounts for heavier uses, and the extent to which heavier vs. lighter uses account for different kinds of outcomes (see also Leinhardt's, 1980 description of implementation studies). Specifically, we can now use the data obtained in the first phase of the evaluation and work with extreme groups, that is - compare only classes (schools, homes) where the use of the programme is the heaviest with those where it is the lightest. The following mapping sentence guides this phase of the evaluation:

$$\text{The extent to which: } \left\{\begin{array}{c} \underline{A} \\ \text{teachers'/parents' attitudes} \\ \text{teachers' training} \\ \text{school/home setting} \\ \text{colleagial atmosphere} \end{array}\right\} \text{ account for: ....}$$

$$\text{variance on: } \left\{\begin{array}{c} \underline{B} \\ \text{little use} \\ \text{moderate use} \\ \text{much use} \end{array}\right\} \text{ and the extent to which: } \left\{\begin{array}{c} \underline{C} \\ \text{facet A alone} \\ \text{facet B alone} \\ \text{facets } A \times B \end{array}\right\}$$

$$\text{account for variance on: } \left\{\begin{array}{c} \underline{D} \\ \text{outcome} \end{array}\right\} \text{ measures with students differing on -}$$

$$\text{measures of: } \left\{\begin{array}{c} \underline{E} \\ \text{prior knowledge} \\ \text{general ability} \\ \text{SES} \end{array}\right\}$$

A number of issues need to be noted in this phase of the evaluation. First, we collect here more detailed (and more costly) data than in the first phase, but we study a far smaller sample. Second, we can examine the extent to which more distal factors (e.g., teachers' expectations; facet A) affect more proximal factors - amount of programme use (facet B), and how each of the two, alone and in interaction (facet C), affect outcomes (facet D) differentially (facet E). Third, as this phase follows the first phase, outcome measures obtained in the first phase serve here as pretest scores. These can either be partialled out to examine the 'net effects' of amount of exposure on outcomes, or used as predictors in interaction with the other variables. Alternatively, this phase can be 'piggy-backed' on the first phase and the outcome data obtained can be used here as well. Fourth, as this phase of the evaluation is based on a relatively small sample, additional outcome measures can be administered, thus increasing the range of observed effects that can be attributed to the distal and proximal variables under investigation. There is however one important factor that the second phase of the evaluation cannot deal with: the contribution of different modes of programme utilization to learning outcomes. The third phase of the evaluation addresses this issue.

The Third Diagnostic Phase: The Contribution of Utilization Modes

Here the idea of the funnel, or cone, is carried one step further. The second phase of the evaluation examines only, say, the upper and lower thirds of the amount-of-use distribution obtained in the first phase. In the present phase we go one step further and select only that third of the distribution where amount of utilization is the highest and examine the extent to which qualitatively different modes of programme utilization, whether at home, in the school, or in both places, affect, alone, or in interaction with other factors, learning outcomes of different learners.

This phase requires systematic observations of how, and in what ways, the programme is utilized, the ways frequent users integrate its materials into curricula, and into classroom activities, the ways programme materials are discussed, the ways students are prepared for viewing, and the like. Johnston and Ettema (1982) have found, for example, in their experiment-like evaluation of Free-style, that when the programme shows were viewed at school accompanied by in-class discussions, learning outcomes were far greater than when the students either watched the programmes in school with no accompanying discussion or watched it voluntarily at home. Such findings would be obtained in the present phase of the evaluation, only that here they follow earlier phases where much additional data are gathered.

The Fourth Phase: Planned Variation

So far, all phases of the evaluation were based on 'naturally provided' variability which, while yielding 'diagnostic' and externally valid data on the conditions that promote or hinder the effectiveness of the programme on a variety of measures, could not possibly provide much internal validity. After all, those who choose to watch the programme more, or those who integrate it more into classroom activities, are not randomly selected; and despite our ability to partial out some background data, the drawbacks of self selection cannot be ignored. Thus, the present phase of the evaluation is based on planned rather than on 'natural' variation. And as this phase comes in the wake of the preceding phases, one is in a better position to choose those variables of interest and importance that merit being manipulated in a relatively small, but well controlled experiment. Such was indeed done with parential encouragement-to-watch in the evaluation of Sesame Street by Ball and Bogatz (1971), by Cook et al (1975) and by Salomon (1977). However, only here can we compare the experimental findings with those obtained under more natural conditions in the preceding phases of the evaluation.

Experimentation has, however, another side to it that pertains to the examination of the programme itself. So far, the evaluation could show the external conditions that facilitate programme effectiveness. But one would still want to 'diagnose' the strong and the weak points of the programme itself. What is the contribution of its humour? Its pictorials? Its structure? What if some of its later units were viewed earlier than initially planned? Or what if some of its contents were read by a narrator rather than dramatized?

Clearly, evaluation of the kinds described above cannot shed much light on these questions unless alternative versions of the programme are produced and experimented with.

The Fifth Phase: Evaluating the Contribution of ETV Attributes

One often wants to know not only what kinds of conditions promote the effectiveness of a single programme, but rather what makes some programmes more effective than others. Such an evaluation could be carried out by means of meta-analytic procedures on the basis of an already existing body of evaluation studies. But meta-analysis of what? Pooling together a number of evaluation studies to examine the overall effective size of, say, amount of exposure would not tell us much about the specific characteristics of ETV production that 'make a difference' beyond the specific contents of programmes.

A possible way of gaining insight into the contribution of specific programme attributes is to treat programmes as we treat individuals who can be given a battery of tests to yield a 'profile' of attributes. Imagine that we had a list of attributes such as 'televisionness', humour, centrality of direct instruction, degree of entertainment, and the like. Each programme (or, if you wish - each single show) can be rated by judges on each of these attributes and be given respective 'grades'.

In our analysis, such 'grades' are entered as independent variables (Akin to intelligence, age, sex, or industriousness of individual subjects) and learners' standardized learning scores - as dependent variables. Now we can ask: How much learning variance across contents is accounted for by, say, humour or by didactiveness? Does effectiveness of such components interact with the specific content of a programme? With the nature of the learner? We would now be able to point to those ETV attributes that do or do not affect learning in any significant way (for a more detailed description of these procedures see Plamer (1974) who describes such an evaluation of Sesame Street by the present author).

Evaluating the Proposed Evaluation

So, we have ETV with a variety of programmes and accompanying materials to be broadcast to schools and homes, and we want to evaluate the merits and worth of the programmes. Are we now in a better position than before? Does the diagnostic evaluation as described and illustrated above handle the complexities involved in ETV evaluation satisfactorily? And is it sufficiently diagnostic to justify its label, the expectations associated with it and its potentially high cost?

The evaluation plan presented above has two major characteristics: the exploitation of diversities, and a funnel-like structure of gradually narrowing foci. The proposed evaluation purposefully employs diversities - of programme features, of audience characteristics, of implementations, of settings, of outcomes - with the explicit intention of finding interrelations among them that predict outcomes. This feature addresses mainly the challenges to ETV evaluation that result from the many diversities and the multi-component, package-like nature of ETV. And it deals with these in ways that permit us to identify bundles of components that facilitate or hinder the attainment of different kinds of outcomes.

The funnel-like structure of the evaluation narrows the focus of attention from a wide angle where everything can interact with everything else, to selected variables of particular interest. This feature addresses the diagnostic intentions of the evaluation in yet another way. It allows us to pinpoint those elements of interest that can be improved in the future and whose improvement might make a difference. It also allows us to identify the kinds of outcomes, whether intended or not, where a particular programme or series have an impact that is more than a mere drop in the ocean.

The value of evaluation, as Cronbach has pointed out, is not in its structure but in its functionality. In the light of the special complexities of ETV, one needs special kinds of evaluation that serve a useful function. It appears that the diagnostic function is worth serving and the way of carrying it out, as suggested here, serves this function satisfactorily.

REFERENCES

Baggett, P 1984. Role of Temporal Overlap of Visual and Auditory Material in Forming Dual Media Associations. Journal of Educational Psychology, 76, 408-417.

Ball, S and Bogatz, G A 1971. The First Year of "Sesame Street": An Evaluation. Princeton, N J, Educational Testing Service.

Bandura, A 1982. Self-Efficacy Mechanism in Human Agency. American Psychologist, 37, 122-148.

Campbell, D T 1969. Reforms as experiments. American Psychologist, 24, 409-429.

Cook, T D et al, 1975. "Sesame Street" Revisited. Russell Sage Foundation.

Cronbach, L J and Associates 1980. Toward Reform of Program Evaluation, Jossey-Bass.

Gombrich, E H 1974. The visual image. D R Olson (ed.), Media and Symbols: The Forms of Expression, Communication, and Education. 73rd Yearbook of the National Society for the Study of Education. Chicago, University of Chicago Press.

Johnston, H and Ettema, J 1982. Positive Images: Breaking Stereotypes with Children's Television. Beverley Hills, CA, Sage.

Kosslyn, S M 1981. The Medium and the Message in Mental Imagery: A Theory. Psychological Review, 88, 46-66.

Leinhardt, G 1980. Modelling and Measuring Educational Treatments in Evaluation. Review of Educational Research, 50, 393-420.

Mayo, J K, Hornik, R C and McAnany, E G 1976. Educational Reform with Television. Standford University Press.

Olson, D R 1974. Introduction. D R Olson (ed.), Media and Symbols: The Forms of Expression, Communication, and Education. 73rd Yearbook of the National Society for the Study of Education. Chicago, University of Chicago Press.

Palmer, E L 1974. Formative research in the production of television for children. D R Olson (ed.), Media and Symbols: The Forms of Expression, Communication, and Education. The University of Chicago Press.

Papert, S 1985. Computer Criticism vs. Technocentric Thinking. Logo 85, Massachusetts Institute of Technology, Cambridge, MA.

Salomon, G 1976. Cognitive Skill Learning Across Cultures. Journal of Communication, 26, 138-145.

Salomon, G 1977. Effects of Encouraging Israeli Mothers to Co-Observe 'Sesame Street' with their Five Year Olds. Child Development, 48, 1146-1151.

Salomon, G 1979. Interaction of Media, Cognition, and Learning. Jossey-Bass.

Salomon, G 1983. The Differential Investment of Mental Effort in Learning from Different Sources. Educational Psychologist, 18, 42-50.

Salomon, G and Leigh, T 1984. Predispositions Debilitate Learning from Television. Journal of Communication, 34, 119-135.

Sanders, J R 1983. The Importance of Context when Studying the Impact of Instructional Television. Paper presented at the Annual Meeting of the Evaluation Research Society, Chicago.

Scarr, S 1985. Constructing Psychology: Making Facts and Fables for our Times. American Psychologist, 40, 499-513.

Schramm, W 1977. Big Media, Little Media. Beverley Hills, Calif, Sage.

Scriven, M 1973. Prologue: Standards for the Evaluation of educational programs and products. G D Borich (ed.). *Evaluation of Educational Programs and Products*. Berkely, Calif, Educational Technology Press.

Shavelson, R and Webb, N 1986. Evaluating Telecourses: Design Considerations and the Concept of Exchangeability. Paper presented at the Annual Meeting of the AERA; San Francisco.

Smith, M S and Bissell, J S 1970. The Impact of Headstart: The Westinghouse-Ohio Head Start Evaluation. *Harvard Educational Review*, 40, 51-105.

Stasz, C et al, 1986. Student Outcomes from the New Literacy Telecourse. Paper presented at the Annual Meeting of the American Educational Association, San Francisco.

CONTINUAL RESPONSE ANALYSIS IN TV EVALUATION

J Baggaley

During 1985, an American sporting goods company announced the invention of the computerised running shoe. Following a run, the shoe is plugged into a home computer, and the runner is provided with immediate feedback on the distance he has covered, the time taken, and the amount of calories burned up. The concept behind the system is sound. Immediate feedback of results can be expected to increase the runner's ability to improve his skills the next time out. He no longer has to rely on intuition in order to gain the maximum return from his athletic efforts.

Similar feedback devices have come on the market for the benefit of film and television producers. Since the development of portable micro-computing facilities in the 1980's, media producers no longer have to rely on questionnaire and interview techniques for information about the impact of their productions. Those techniques were in any case largely unable to provide the specific information which producers require about the impact of particular production techniques. Precise moment-by-moment feedback of audience reactions to a production, is now available, generated by a wide range of electronic techniques.

Via the new systems, the audience's responses can be recorded continuously throughout the production, and fluctuations within them instantly analysed. When the continuous record of audience responses is synchronized with the production itself, the producer can inspect the momentary fluctuations in response which are associated with individual scenes and production techniques in time to re-shoot or re-edit the programme for greater effect.

The following chapter discusses the benefits of these innovative evaluation techniques, and methodological problems which commonly arise when they are used. The new perspectives provided by continual response analysis in 'formative evaluation' are discussed, and uses of the technique are indicated in a variety of research contexts. The chapter also discusses the political problems which media researchers and producers commonly face in the evaluation situation, and suggests that, sensitively used, the new techniques can create a more productive relationship between them.

Audience Response Technologies

The principle of continual response analysis is by no means new. In the 1870's the French physiologist Jean Ettienne Marey developed mechanical systems to measure relationships between psychological responses and heart-beat rates. Electronic systems for monitoring the fluctuations in audience reactions have been used in commercial broadcasting since the 1940's. Lazarsfeld and Stanton (1944) developed a mechanical 'Program Analyser' to evaluate the impact of radio programming. The Analyser provided moment-by-moment feedback from several hundred listeners simultaneously, and it was used in academic and military contexts as well as in broadcasting research. Similar systems were the Reactograph, designed by Cirlin and Peterman (1947), and the Film Analyser of Carpenter et al (1950). Between 1952-58, a fully equipped laboratory for the analysis of psycho-physiological responses to films was built in Brussels by the researcher Sallesi Lavergne. The history of these biometric research systems is discussed by Cambre (1981) and Malik (1981).

During the 1970's, a range of computer-based audience response technologies became available. Facilities for measuring responses to a lecture presentation are a common feature of North American college lecture theatres.

Via a set of electronic buttons built into their chairs, students may answer true/false, multiple-choice, or opinion-type questions at the lecturer's bidding. An immediate summary of the responses can be presented on analogue or digital displays at the front of the room. In this way, the student receives feedback about his peers' responses, and the lecturer gains information about the group's attitudes or achievement.

Numerous systems of this sort have been developed in clinical and social psychological research settings (Clarke and Ellgring, 1983). Several investigators have reported uses of the Datamyte system to record patterns of behaviour in human and non-human social groups (Conger and McLeod, 1977: Sawin et al, 1977; Scott and Masi, 1977; Sykes, 1977). A hand-held response unit stores an observer's responses to the behaviour, and transmits it to a computer for analysis and feedback about twenty minutes later. Changes in, for example, the frequency, average length and variability of behaviour can then be examined across time. This particular system is relatively expensive, however, for the one hand-unit provided.

Response analysis systems differ on a large number of bases in addition to cost: notably portability and flexibility, speed and level of analysis, and clarity of feedback. In broadcasting research, of course, the common need is for multiple hand-units to record the responses of a whole audience. The Multi-term response system fulfils this need via a hard-wired network of up to 250 response units. The Program Evaluation Analysis Computer (PEAC system) collects the responses of a potentially infinite number of people via a set of remote battery-powered units. With the latter facility, reactions to a production may be collected simultaneously in a range of settings (e.g. viewers' homes), as well as in a central location. Comparable methodologies include the CBS Viewer system, the technique of Audience Studies Inc, the PAL Consumer's Computer, the French Mediascope, and the Percy (Voxbox) system.

An extreme form of the technology has been developed by Warner Communications, and is known as the QUBE system. The system is used to monitor audience reactions during TV-viewing, and involves over 30 000 audience response units situated in homes in Columbus, Ohio. The units permit instantaneous, two-way interaction between the viewers and the QUBE computer, and have been used to obtain 'instant polling' of opinions on political and social issues, and to allow viewers to select, for example, alternative endings to TV drama productions. The system has commonly been criticised for the non-scientific way in which it may be used to shape public opinion.

All forms of continual response analysis can be subjected to this type of criticism. The speed and clarity of data provided by the electronic response analysers in programme pilot-testing are beyond question. To the designers of a TV or film production, of course, feedback about its moment-by-moment appeal can be irresistable. Programme segments can be adjusted or extended and camera angles altered - even during live presentations - in order to maintain and enhance programme appeal. However, the quality of such feedback is only as sophisticated as the research methods which were used to generate it; and impulsive interpretations of hastily gathered data can be highly suspect.

In educational broadcasting, for instance, the extent to which a measure such as moment-by-moment appeal can actually predict overall learning is debatable. Similarly, little is known about the criteria by which a meaningful shift in response can be distinguished from a random one. Many media producers are rightly defensive about the introduction of continual response methods into their industry for such reasons. They suspect that audience researchers will use the methods to dictate aspects of production content on quite unjustified bases. Urgent attention must therefore be paid to the research methodology on which such systems depend, while the broadcasters are still willing to consider their benefits.

A comparison between three of the leading methodologies in North American broadcasting research was made by the Corporation for Public Broadcasting (1981).

Programme pilot-test results obtained by two of the electronic methods (PEAC and Percy) were compared with those of a more conventional testing method, the discussion or 'focus' group. The three approaches were judged in terms of a) response articulateness versus objectivity; b) sampling flexibility; c) practical benefits to programme producers; and d) long-term benefits to programme policy-makers and distributors.

The conclusions of the CPB study may be summarized as follows. While the openness of the focus-group situation usually allows discussants to be relatively flexible and uninhibited in their responses, it can also have inhibiting effects. Powerful group biases can affect the opinions expressed. The opinions of individual group members may be dominated by those of more assertive individuals. By the time the presentation is over and the discussion takes place, viewers may also have forgotten many of the fleeting but critical reactions they experienced whilst the presentation was in progress. The ability to record one's responses to a programme simultaneously, via a hand-held response unit, can reduce these problems. Being non-verbal, responses are usually private and anonymous. Audience members have the opportunity to make a completely uninhibited assessment, and to change it as frequently as they choose. On the other hand, an automated response task invariably restricts the range of available responses to a set of fixed options. The CPB study (1981) concluded that the most effective testing situation for the foreseeable future will probably be one featuring the electronic and focus-group methodologies simultaneously.

Certainly, the electronic techniques are the only current means whereby moment-by-moment fluctuations in audience impact can be measured effectively: one would hesitate to stop the programme every few seconds for a discussion! The imposition of a closed-ended response can be seen as a worthwhile price to pay for this extra information. On the other hand, the overall impact of a programme is unlikely to be established other than by post-test measures (e.g. questionnaire or discussion methods). The unique benefits of continual response analysis are its ability to show the relationship between minute fluctuations in audience response and momentary, critical programme events.

Given the appropriate analysis software, the moment-by-moment responses may also be examined for individual differences between viewers, and insights gained into the reactions of different types of viewer to particular production techniques. In this way insights can be gained into the effects of, for example, programme pacing and illustration techniques upon viewers of different ages or abilities. For programme policy-makers, continual response analysis can answer general questions concerning, for example, the reactions of different audience types to programme violence and stereotyping.

The uses and limitations of continual response analysis methods are illustrated in the following case-studies. The studies were conducted by the author between 1980 and 1984, initially at Memorial University of Newfoundland, and more recently at Concordia University, Montreal. Both universities had purchased, for media research purposes, the Programme Evaluation Analysis Computer. The PEAC system was developed in the late 1970's by researchers at TVOntario in Toronto and at the Children's TV Workshop, New York (Nickerson, 1979). It was selected from the range of options discussed above on the basis of its superior portability and flexibility of operation, and its relative cost-efficiency.

'Instant Polling' of Audience Responses

In November 1980, the PEAC system was used to assess public reactions to the televised debate between American President Jimmy Carter and the presidential challenger Ronald Reagan. In St. John's, Newfoundland, a panel of two dozen viewers watched the debate in their homes. As they did so, they used the portable PEAC hand-units to respond to the following question: "Who, from one moment to the next, is winning the most votes?". Three options were available to them, on buttons labelled CARTER, REAGAN, and DON'T KNOW. Their continual responses were sampled at 4-second intervals. The study was conducted on behalf of the Canadian Broadcasting Corporation (CBC) as material for a radio report on the debate the following morning.

The results of the Carter-Reagan study have been reported by Baggaley (1985a). Although limited in its scope and generalizability, the study gave indications of the telling impact of non-verbal strategy in the debate, and of the speed with which the contender Reagan was able to dominate President Carter in the viewers' eyes. A specific tendency was noted for viewers to react adversely to Carter at moments when he was presented in a particular side-shot. It was concluded that the President appeared more tense from this camera angle than he did from others; and the conclusion was reported on CBC-Radio the following morning. The observation was also reported eighteen months later by Wingerson (1982).

The benefits of precise and immediate feedback of this type are three-fold. Clearly, by being forewarned, the public may become forearmed about the deliberate use of debating strategies in future media campaigns. The producers may also learn from the feedback, by avoiding the possibilities of such bias. More insidiously, such information can provide the politicians themselves with the means to refine and enhance their campaigning strategies. Indeed, the PEAC technique - virtually unheard of in 1980 - has since been reported as used during the 1984 Reagan-Mondale debates by the Reagan campaign team itself (Toronto Globe and Mail, 1985).

The rates of audience response during the first eight minutes of the Carter-Reagan debate are plotted graphically in Figures 1. The four graphs are divided into segments, according to the alternating question-and-answer format of the debate. The first four minutes is dominated by responses on the DON'T KNOW button (Fig. 1a) and the CARTER button (Fig. 1b). Viewers perceived the incumbent Carter as winning the most votes even before he spoke. During the fifth minute, however, votes began to accumulate for Reagan (Fig. 1c), and in the subsequent course of the debate, Reagan's perceived 'voteability' increased dramatically, particularly in the nineteenth minute during his discussion of Carter's economic record. The peaks of response on the three buttons may be compared in the combined graph (Fig. 1d).

Figure 1(a) : DON'T KNOW

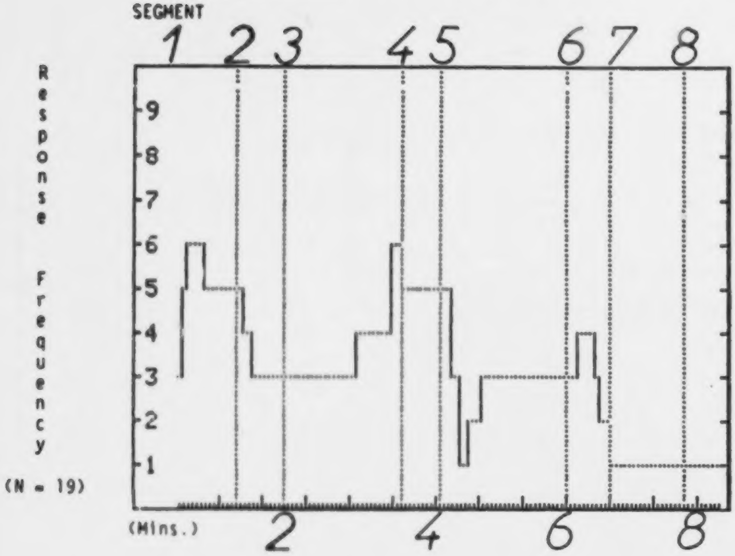


Figure 1(b) : CARTER

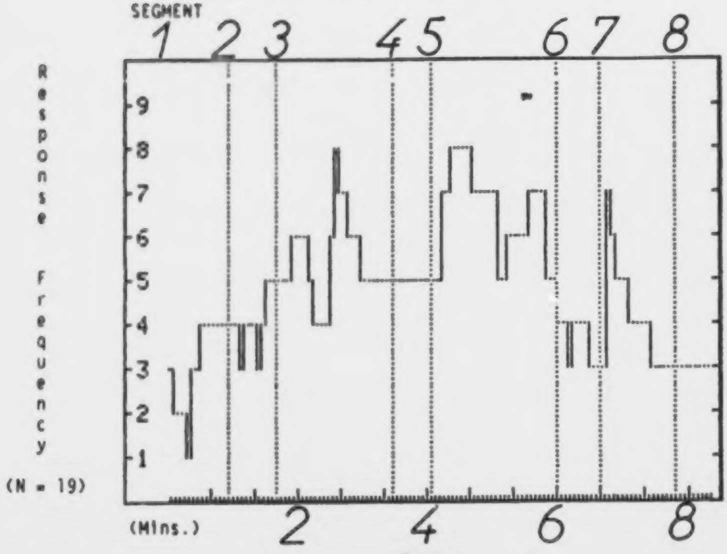


Figure 1(c) : REAGAN

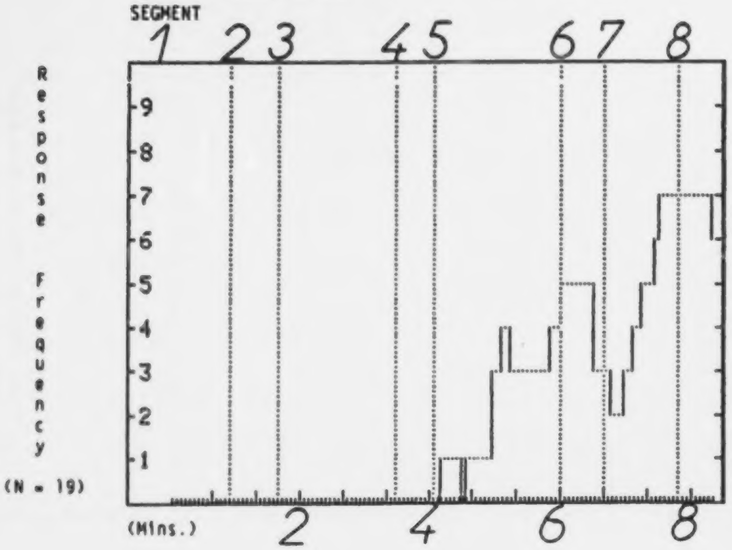
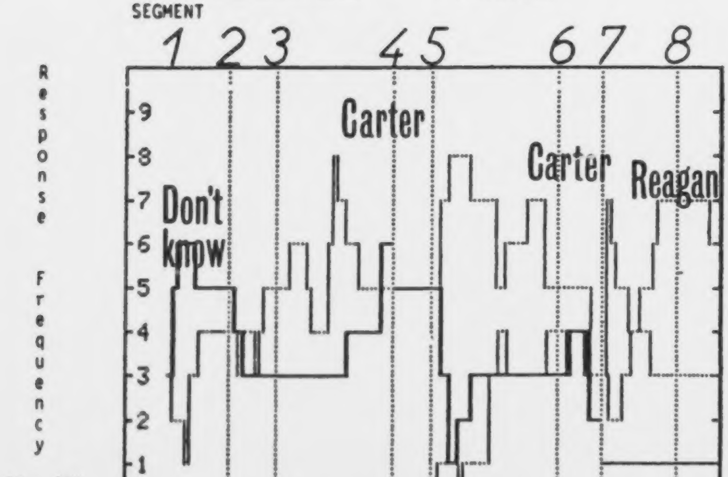


Figure 1(d) : COMBINED



The problems of inferring overall impact from such data are obvious. Firstly, the panel of subjects used in the Newfoundland study was minimal in size, and any attempt to generalize from their responses to the larger American audience would be highly questionable. (One can sympathise with the accused in a court of law, for whom life and death depend on the reactions of a jury half this size). If such a study is to be beyond reproach, therefore, it must clearly make use of a representative and balanced sample of the audience for whom the programme is intended.

Continual response studies must also involve a response task within the intellectual and physical means of the subjects. As the following section shows, typical response tasks can often be too difficult for some viewers to handle.

Complexity of the Continual Response Task

In common with other response analysers, the PEAC system allows for the collection of continual responses on one of two bases:

- nominal, categorical responses - such as CARTER, REAGAN, or DON'T KNOW; or
- an interval scale of responses - such as GOOD, FAIRLY GOOD, FAIRLY POOR, POOR.

It also provides two modes of push-button response: the Reset mode, in which the appropriate button must be depressed continuously in order for a response to be registered, and the Latched mode in which the current response is assumed to pertain - whether the appropriate button is currently being pressed or not - until a different button is pressed. The Reset mode can be useful in situations where subjects are required to give a fast rate of response over relatively short periods, possibly in relation to specific presentation variables or cues. The Latched mode can be useful in non-cued, more leisurely response situations. Instructions to respondents prior to testing have to bear these decisions carefully in mind.

A further decision in designing the response task concerns whether the respondents are to be forced to express an opinion at every moment of the presentation, or can be free to signify momentary indecision or indifference towards it. When an interval scale is in use, respondents are usually happy to give at least a tentative response (e.g. FAIRLY GOOD or FAIRLY POOR). However, in situations where a fast response rate and/or numerous response options are involved, respondents may be allowed the right to a momentary rest and the absence of any overt opinion at all. In the Latched mode, this can be provided via a DON'T KNOW or NO OPINION button. In the Reset mode, it can be signified simply by an absence of response on any button.

The response task actually selected will usually depend on the nature of the research question at hand, and on the respondents' physical and mental abilities. The simple nominal response task (CARTER or REAGAN, for example) presents relatively few problems. The respondents record their fluctuating opinions first on one button and then on another as appropriate throughout the programme. The number of response options depends primarily on the respondents' ability to select between them at the requisite speed. When the PEAC system is used, a change of response can be recorded as fast as every quarter-second - a rate which is capable of measuring the attitude changes of even the most fickle of television viewers!

An interval response task, however, is more intellectually complex. Using an interval scale, the respondent must not only inspect continually his current state of response, but he must relate the scalar magnitude of his current response to that of his previous response. Given the additional responsibility of attending to the media presentation itself, a task involving more than one interval scale is usually out of the question. For some respondents, and over long periods, the intellectual demands of even one continuous interval task can be excessive; and if the hand-unit buttons have to be continuously depressed in order for the response to be registered, the physical demands of the task may be excessive over even short periods.

(N.B.: The response units used with the PEAC system are incapable of storing simultaneous responses by a viewer on more than one continuous interval scale, whether he is capable of making them or not.)

In a study reported by Baggaley and Smith (1982) the continual responses of Newfoundland fisherman to a film on the Canadian seal harvest were investigated. The film was entitled 'A-1'. When the men were asked to assess it on a 4-point interval scale (from GOOD to POOR), they tended to use the two extreme buttons only. Of these two buttons, the POOR option was much less frequently used than the GOOD option. The Latched response mode was in operation, and accordingly the respondents were instructed that successive responses on any one button would not be necessary. However, their uses of the GOOD button were characterised by repeated responses of this type. It appeared that the men's usage of the buttons had for the most part been dictated by a desire to signify momentary, positive reactions only. At all other moments they appeared to be either impartial about the film, or unwilling to register a negative reaction.

A second group of fishermen was asked to use the GOOD and POOR buttons only. The Reset mode was used, and the men were instructed to maintain the pressure on each button until the response was no longer appropriate. This they did, sometimes for minutes on end, with no problem. The researchers had learned that some subjects may find a 4-point interval response task too complex, and to anticipate the mental and physical demands of each testing situation.

In an investigation by Baggaley (1985b), 96 children from 3-6 years of age reacted to an eight-minute video sequence of animated cartoon and puppet characters. The same response task was used as in the previous study - a Reset response mode, with the instruction to keep each button pressed for as long the response was appropriate. The children were cued to indicate their relative enjoyment of each successive TV character via a SMILE or FROWN button (Figures 2). Figure 2a presents the frequencies of SMILE responses at 2-second intervals during the first four minutes of the presentation. Figure 2b indicates the FROWN and 'no response' rates for the same period.

Figure 2(a) : SMILE

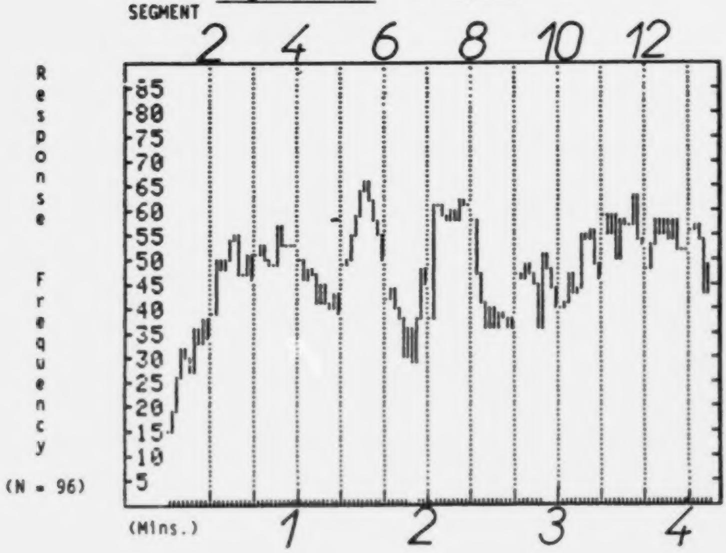
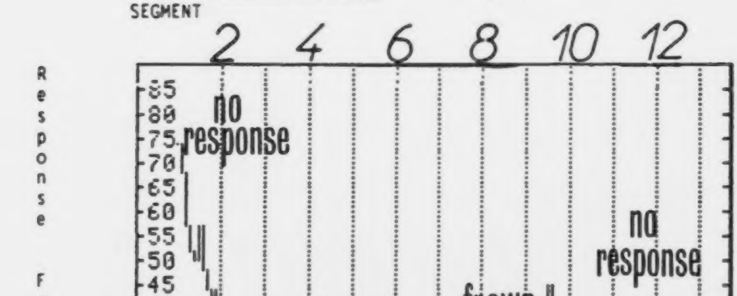


Figure 2(b) : FROWN



The consistency of responses to each character in this study remained high across a series of repeated presentations. Though the reliability of the 3-year olds' responses was questionable, the 4-6 year olds adjusted to the task with ease - somewhat to the surprise of their Kindergarten teachers! It appeared that the push-button technology had given the children an opportunity to express levels of critical assessment that their powers of language as yet did not.

In attempting to design a response task which is within the respondents' mental and physical capacities, therefore, the following decisions are recommended:

- *The number of variables to be investigated simultaneously should be decided according to the respondents' customary ability to cope with them. In the absence of an exact precedent, the task should be pilot-tested with respondents representing the audience in question. If one variable only is to be assessed (e.g. the moment-by-moment credibility of President Carter) an interval scale may be considered. If several variables are to be assessed (e.g. the merits of Carter and Reagan simultaneously) separate nominal measures should be used. If the selected measure subsequently proves too complex for the respondents, the task can be simplified either by converting the measure from an interval to a nominal level, or by reducing the number of variables under scrutiny.*

- *The decision to force the response or to allow an undecided response should also be taken on the basis of the subjects' likely response capabilities. If an interval scale is in use, an 'Undecided' mid-point button may be provided. If nominal responses are involved, a 'no response' button may be provided, actively cancelling the responses made via other buttons. In either situation, the Reset mode may be employed, allowing the respondent to register an absence of response by simply not responding!*

- *The extent to which subjects can be cued to respond to specific variables in the presentation, or at particular rates, should be decided primarily on the basis of the length of the presentation. At present, the tolerance of respondents for tasks of different lengths can only be judged intuitively. When the feasibility of a cued response task is in doubt, it should be avoided or the presentation shortened.*

- *In all normal test situations, it is desirable for each respondent to be able to refer to a visual display of his most recent response on the hand-unit (as via the PEAC system). In this way, the respondent is reinforced in his efforts, and he may also be reminded of his last response, as in Latched and interval situations.*

Interpretation of Continual Response Data

The data provided by continual response technologies can be both graphic and beguiling. Peaks and troughs in the moment-by-moment response profile invite instant interpretations of, for example, 'high visual interest', 'medium programme appeal', 'low presenter credibility' - and so on, depending on the response measure used. Such interpretations may be quite invalid. In one case, the writer had to restrain a TV producer from summarily firing the programme presenter in response to low rates of audience reaction that were observed during his appearances. It was pointed out that a low rating for visual appeal did not necessarily disqualify the presenter as a good educator. Conversely, a programme or programme presenter may receive a consistently high moment-by-moment rating, and yet be obviously failing in its/his attempt to fulfil the main programme objective.

Considered in isolation, the predictive validity of continual response data is actually very low. In common with other forms of data gathered sequentially, they are subject to various types of psychometric error. When the continual ratings of a programme are generally positive, for instance, a momentary lapse in quality may well be under-estimated by viewers. When a programme segment follows one which is highly unpopular, on the other hand, its momentary rating may be over-estimated. These tendencies to over- or under-estimate in a continual response task are identified as 'series' and 'time-order' effects respectively (Woodworth and Schlosberg, 1961).

Moreover, in analysing interval data particularly, one cannot always be sure whether a momentary response is more appropriately interpreted as an absolute judgement on the scale, or as a relative one. A viewer's response on the FAIRLY GOOD button, for example, may be construed at its face value as representing an absolute judgement of 'fairly good'. But it might also represent a sudden, immense improvement in perceived quality from VERY POOR to FAIRLY GOOD, and a relative judgement whose correct interpretation is identical to that of a shift from FAIRLY POOR to EXCELLENT. The viewer's current choice of buttons in a continual response task is dependent on relative as well as absolute judgemental forces in this manner. When momentary responses are averaged across a group of viewers, their precise psychological meaning becomes even less apparent.

Clearly, the attempt must be made to validate continual response data by refering them to general criteria for programme effectiveness. The criterion-referencing of continual response data can usually be achieved by -

- comparisons between the responses of different viewing groups; or

- comparisons between moment-by-moment responses and a measure of overall programme impact as yielded by a pre- and/or post-test.

The comparison of responses between groups may be appropriate in situations where a producer requires evidence of the programme elements which are capable of interesting one particular audience sub-group as opposed to others. For example, in the study by Baggaley (1985b) of pre-school children's responses to TV cartoon and puppet characters (see previous section), particular comparisons were made between the reactions of the boys and girls, and between those of English- and French-speaking children. When the continual responses of the boys were compared with those of the girls, gender differences in their preferences for particular characters emerged. No such difference was observed on the basis of the children's cultural background. The sponsors of the study - the National Film Board of Canada - received feedback about the types of TV character most likely to appeal to both boys and girls simultaneously.

In the study reported by Baggaley and Smith (1982), on the other hand, a continuous measure of audience response was compared with measures of overall programme impact derived from pre- and post-tests. The film in question concerned the Canadian seal-hunt, a controversial object of protest by international conservationist movements. The continuous measure of response was one of general approval towards the film, on a scale from GOOD to POOR. The overall measures related to shifts in attitude towards the seal-hunt, and in learning about it, as measured from immediately before the film to immediately after it. If a positive continual response in such a situation were to be accompanied by a negative overall impact on attitudes or learning, it would be obvious that the overall responses had greater validity as an educational index. The high continual responses would be either 'not high enough' in relation to the overall effect; or they could actually be quite irrelevant to it. Only when used as complementary to overall criteria, can continual response data have predictive meaning.

Particularly vital information in this study was gained from the responses of a group of Newfoundland high-school students. At first glance, their data seemed to indicate the type of disapproval shown towards the film by the seal-hunt protesters.

However, on closer examination of their data, a totally contrasting interpretation was found to be tenable. At a particular moment in the film, the killing of a seal was shown. At the same moment, a sudden shift was observed in the audience's continual responses towards the negative end of the approval scale. When a nominal response scale is used, response fluctuations of this type are apparent in terms of the number of audience members shifting from one button to another at a given moment, as in Figures 1 and 2. When a series of buttons representing an interval scale has been used, the levels of response may be assessed in terms of either

a) *the frequencies of response on each button individually, or of*

b) *the average response on all of the buttons at once.*

The shift towards disapproval by the high-school students on seeing the seal killed is apparent in Figure 3a. The graph shows a 50 percent drop in the number of students pressing the GOOD button at that particular moment in the film (i.e. during the eighth minute). The Figure may be compared with Figure 3b, in which the group's average response on all four of the interval-scale buttons is presented - a more precise profile based on far more information. Both graphs are, of course, totally ambiguous with regard to the meaning of any particular moment-by-moment response. In this case, the researchers decided that the sudden response shift signified either distaste for the killing of seals, or disapproval of the film for showing it, or a combination of both. In an attempt to determine which of these hypotheses was the most tenable, the responses at this moment in the film were referred to the information about each respondent collected via the pre- and post-tests.

Figure 3(a) : GOOD

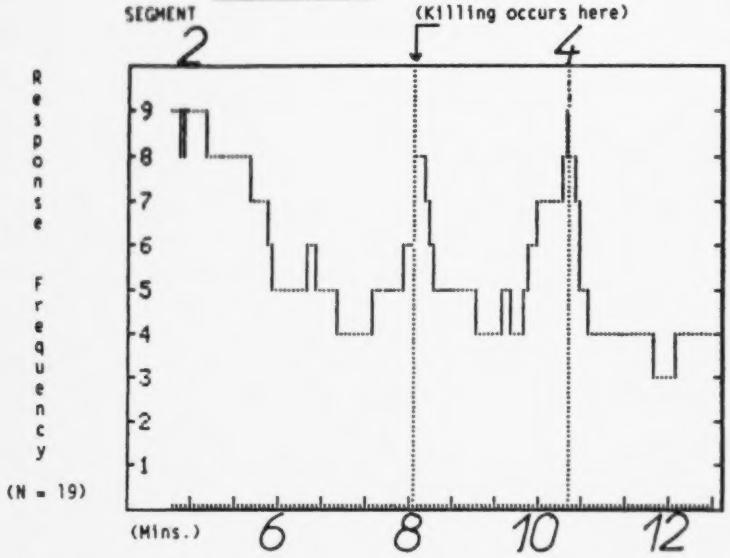
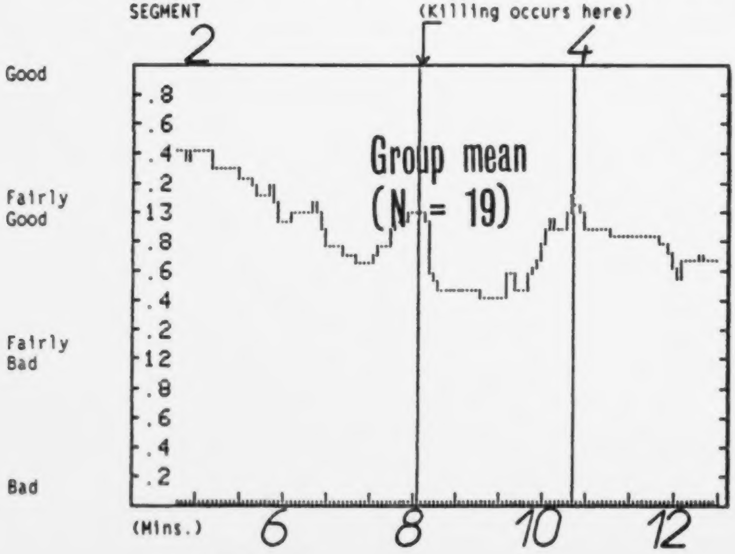


Figure 3(b) : ALL FOUR INTERVAL-SCALE BUTTONS



Individual differences in responses to the killing sequence could thus be related -

- *to demographic information about the respondents - their age, sex, family background, etc, as well as*
- *to their prejudices about the seal-hunt, and*
- *to changes in their attitudes after seeing the film.*

A marked difference was observed in the responses to the killing sequence given by the boys and the girls. Figure 3c indicates the particular disapproval shown by the girls in the group; at first, the researchers interpreted this as indicating a greater level of 'squeamishness' by the girls towards the killing scene. However, a closer inspection of the demographic data revealed that most of the girls in the audience were from seal-fishing families, whereas the boys were not. It was doubtful that the girls could be raised in a seal-fishing environment and be more squeamish about the act of harvesting than those who were not. The remaining analysis discounted this possibility.

Figure 3d isolates the response to the killing sequence of the girls whose fathers were seal-fisherman. It is apparent that the negative response to this scene is due to the virtually unanimous disapproval expressed by these particular respondents. Inspection of the overall reactions to the film, however, indicated that children from sealing families were quite used to the killing of seals, and approved of it as a major source of their families' income. They would not, however, approve of the general public's seeing the film, for fear that their fathers would be viewed in a barbarous light. Rather than connoting a distaste for the seal-hunt, therefore, the continual response data were eventually interpreted as revealing disapproval of the film at that particular point, borne of a distinctly pro-sealing attitude.

Figure 3(c) : RESPONSES BY GIRLS

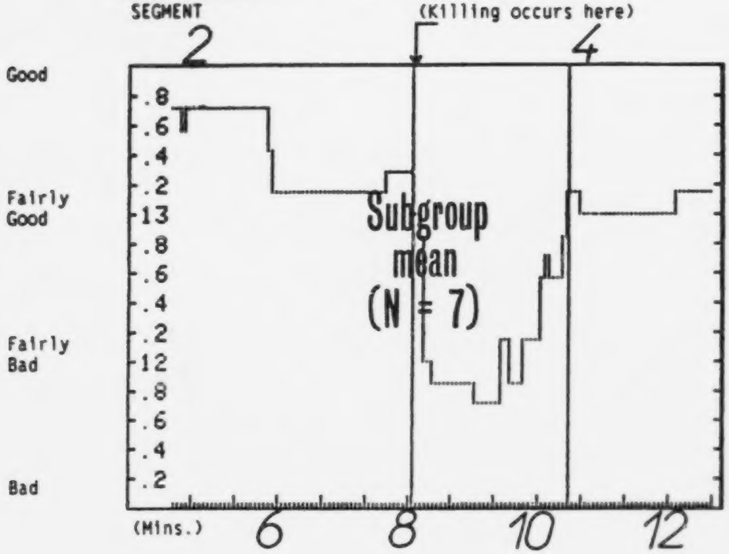
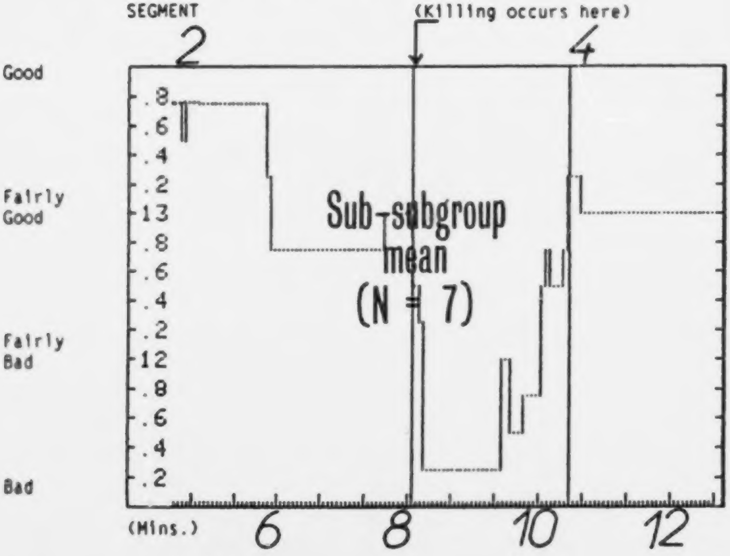


Figure 3(d) : RESPONSES BY GIRLS WHOSE FATHERS WERE SEAL-FISHERMAN



Clearly, the repertoire of data gathered and sifted by computer-based techniques of this sort can generate a series of quite elaborate hypotheses for testing. They do so with speed as well as finesse. The above hypotheses about reactions to a particular film sequence were formed and modified by the researcher and the producer of the film in the space of less than 10 minutes in a motel-room analysis session following the rural test session. The conclusions were confirmed the following morning in a debriefing session with the respondents. From the same study, the producer gained evidence about weaknesses in some of his visual illustrations, and about the valuable role played by his soundtrack music in sustaining audience interest. The latter piece of evidence proved useful in persuading the government sponsors of the film that such production values should be retained in the film, having educational value (Baggaley and Smith, 1982).

The very speed with which meaning can be extracted from such data, however, can distract from their inherent weaknesses. In the study of the sealing film, it is very clear - as in the Carter-Reagan study discussed earlier - that the small number of respondents prevents us from drawing specific conclusions with the degree of certainty that would be expected in a more rigorous research study. When discussing the specific sub-subgroup of respondents belonging to the select coterie of girls from seal-fishing families, we are reduced to drawing conclusions on the basis of four persons only (Fig. 3d)! The lack of sufficient audience numbers for inferential purposes is the programme evaluator's most common problem. Unfortunately, many of the most fundamental determinants of the impact of a programme do not become apparent until the evaluation study has been completed. With hindsight, it becomes clear that the sampling or respondents in the study was too narrow to prove or to refute the hypotheses now apparent. In many evaluation situations, shortages of both time and money prevent the collection of further data, and the evaluator is reduced to offering tantilising, though ill-substantiated trifles.

The writer's practice in these situations is to present the available evidence to the film producer, free of interpretation, and to let him draw his own conclusions. If the researcher is dogmatic in presenting his own conclusions and recommendations about a production, his relationship with the producer will probably flounder in any case. If, on the other hand, the producer is happy to form his own hypotheses from the evaluation evidence - however tenuous they may be - the evaluator has succeeded in fulfilling the most vital of his objectives, which is to provide at least modest assistance in the production process. The producer will be able to test his hypotheses in practice - perhaps by shortening a sequence or by altering its overall position - and can reject them as he sees fit. In the past history of media research, an outcome of this sort would have been considered a triumphal indication of the evaluation process.

The evaluator inevitably faces problems of reliability and generalizability whenever - as is normally the case - the product he is inspecting involves any element of novelty. Faced by the wholly unpredictable reactions of an audience he has not previously encountered, to subject matter and production techniques which have not been previously combined, he must make the most judicious sampling of techniques and human feedback available to him. The same choice is open to the producer. The only advice that can be offered to either of them is to be pragmatic in approach and cautious in interpretation.

Formative Evaluation and Research

As the volume of literature on continual response analysis increases, it is likely that an increasing amount of information will be capable of transfer from one study to another. From the study of responses to the seal-fishing film, for example (Baggaley and Smith, 1982), it was possible to generalise about the techniques to be used in producing other films for rural audiences.

It was clear, for instance, that the fishermen responded enthusiastically to scenes showing familiar people, or familiar types of people, places and equipment. Via repeatedly showing such scenes, the film sustained the men's interest and unltimately proved most instructive for them. It may be assumed that the careful inclusion of familiar elements would enhance similarly the educational effectiveness of other films.

If the good sense behind this practice seems obvious, it should be contrasted with the production techniques used in many conventional instructional films. International agencies spend millions of dollars annually on film and television productions which are blithely assumed to meet the needs of their intended audiences. Lavish productions concerning health and work habits are released for a wide range of audiences, both educated and lesser educated. Rural audiences are constantly expected to identify with films centering around the unfamiliar activities and types of people found in urban communities; and the attention paid to pre- and pilot-testing of the film's educational impact is minimal.

As the availability of funds for the production of educational media materials decreases, so the need for evaluation and improvement of their cost-effectiveness is intensified. In 1981, one national health agency was sufficiently concerned with the effectiveness of its educational media materials, that it embarked upon a detailed evaluation study of their impact upon a wide ranging sample of the intended audience. The Canadian Cancer Society (CCS) had produced and distributed a wide range of cancer education films for all sectors of the Canadian public. It suspected that many of its educational materials - particularly those reliant on reading skills - were providing little or no benefit to the sectors of society with the most need for them. In areas such as lung and skin cancer prevention, the most needy sectors were perceived as the rural and 'functionally illiterate' communities.

The DEMO Project (Baggaley, 1985c) was designed to investigate this possibility, and to recommend ways in which the impact of the CCS's public education programme might be improved. By evaluating the impact of specific films, the project aimed to derive generalizable research conclusions in this regard. In fact, the need for improvement of the materials was found to be more severe than had been initially assumed. Male audiences generally were found to be defensive on cancer education matters. Their resistance to the types of film currently available to them transcended educational and social boundaries. The viewers receiving the least benefit from conventional smoking prevention films were those who smoke. Only the non-smokers were reinforced in the belief that smoking is unpleasant and should be avoided. After viewing some of the films - professionally produced by leading Canadian and American production houses - the smokers in the audience were more militant about their right to smoke than they had been beforehand. Figures 4 reflect the responses of 93 people to one such film ('Smoking: the Unconscious Act'). In Figure 4a, a steadily increasing rate of approval is indicated by the average responses of the sample from one moment to the next on a 4-point scale. Particular segments of the film are seen as more or less effective on this basis. The usual problems of interpreting the graph are faced, of course. Although a high-point in the film's perceived value is evident at the beginning of the eleventh minute, there is no means of determining from the graph whether that moment is critical to the film's overall impact. The viewers' responses in general appear positive towards the film from the eleventh minute onward, but one cannot tell from the graph alone whether they are 'positive enough'. Once the continual response data were related to the independent demographic and attitudinal data, however, the meaning of the graph became gradually apparent. Breakdowns of the continual responses according to independent demographic data indicated few significant differences based on such factors as sex or age. The one variable which did affect the continual responses, however, was the audience's smoking behaviour; and in this case, fortunately, a statistically sufficient number of respondents was available for the purposes of a reliable analysis.

Figure 4(a)

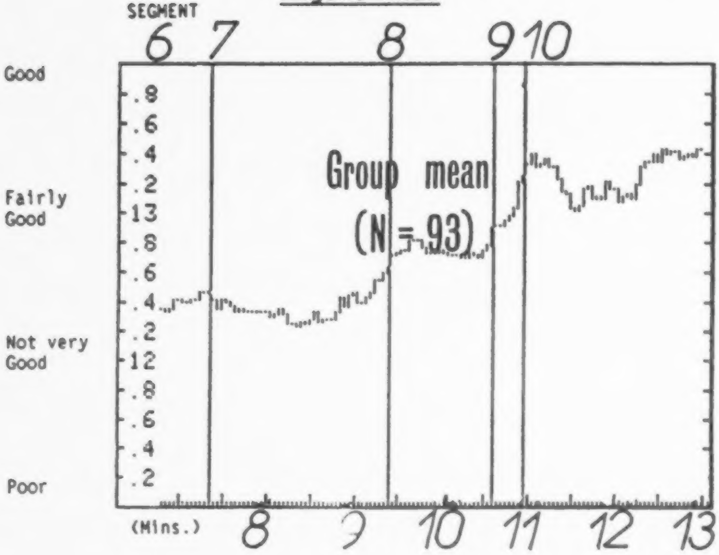


Figure 4(b)

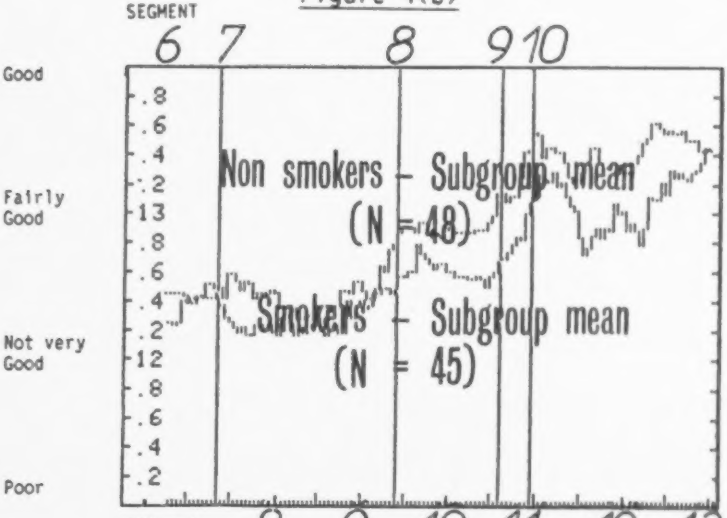


Figure 4b shows that the respondents who smoke were substantially more negative about 'Smoking: the Unconscious Act' from an early stage than were those who did not. If the continual responses of the non-smokers were indeed 'positive enough', those of the target audience of smokers were probably not. Particularly interesting are the moments at which the smokers and non-smokers' responses were seen to diverge - at the beginning of segment #7, for instance, and in the second half of the eleventh minute. For these are the moments at which the perceptions of smokers and non-smokers to the film can be differentiated.

By synchronising the continual and demographic data with the film itself, the researchers were able to identify specific techniques consistently associated with defensive overall reactions on the smokers' part (Baggaley, 1986a). These conclusions were reinforced by accompanying studies of three other smoking prevention films. Smokers were found to be extremely sensitive to any suggestion, however unintentional, that they were being 'preached to' about smoking. Their responses became more negative at moments in the films when visual symbols of medical authority appear (such as a white surgical coat or a lung X-ray), and they appeared to spend much of their time anticipating a didactic message. They were also apt to give increasingly adverse reactions at moments which suggested, however unwittingly, that smoking cessation could make them appear socially eccentric.

Recommendations regarding the types of scenario to be avoided in future films on this topic were made to the CCS on this basis.

On the positive side, smokers' responses became more favourable at moments when practical tips for smoking cessation were given. They seemed to appreciate acknowledgements that smoking is hard to give up, though they did not approve of suggestions that sometimes it can be too hard. The final evaluation report suggested that a fine line has to be drawn between optimism and pessimism on lung cancer prevention.

Such films should maintain, it was suggested, a healthy degree of fear on the smokers' part regarding the dangers of smoking and the prospects for cancer cure; they should however, increase optimism regarding the prospects for cancer prevention.

These conclusions were later applied in the design of a new CCS film on skin cancer prevention (Baggaley, 1986b). The film presented its designers with a particular challenge, in that it aimed to convey an unpopular message (self-protection from the sun) to a wide-ranging audience. Particular emphasis was to be placed on the dangers of skin cancer faced by rural males - a sub-group found to be highly resistant to cancer prevention messages during the earlier research. To meet this challenge the CCS commissioned a series of research studies conducted before, during, and after the production process. Each of the studies employed continual response analysis by means of the PEAC system.

Prior to the production of the film, the responses of 250 urban and rural viewers to an earlier film on skin cancer were analysed. The moments of relative strength and weakness in the old film were identified, and a report was submitted to the producer of the new film concerning the techniques likely to be most successful with different audiences. It was recommended, for example, that the types of information required by the urban audience should be conveyed in segments using fast-paced camerawork and editing techniques. The segments intended for rural viewers should be motivational rather than packed with facts, filmed in a rural context, and presented at a gentler cutting rate. The rural segments might also feature a doctor character, in view of the interest shown by rural viewers in an authoritative doctor figure during the older film.

When the new film reached the rough-cut stage, it was pilot-tested on a new sample of 128 viewers, and its moments of strength and weakness were identified using the same research methods.

The film producer responded to these results by shortening various segments, by modifying the voiceover to strengthen character identification, adding montage and graphic sequences to emphasise some of the educational points, repeating certain sequences in order to create relief, and adding music to heighten mood and structure. The film was then post-tested on a further sample of 194 viewers, and the impact of the pilot and modified versions was compared.

Figures 5 indicate the contrast between the momentary responses of a typical audience sub-group to the pilot and modified versions. Whereas the pilot version had major weaknesses in at least four sequences (Fig. 5a), the modified version appears strong throughout (Fig. 5b). Adverse reactions to the four weak sequences are either reduced or eliminated altogether. The film now seems to be highly successful both in its ability to sustain the interest of a wide range of viewers from moment to moment, and in its overall impact. Its designers acknowledge that the research process has generated a film quite different from that which would have been created otherwise; for it has uncovered numerous quirks of audience response which would not have been foreseen without it (Baggaley, 1986b).

Each of the studies cited during this chapter indicates the unique role that continual response analysis can play in the development of new media products. The second-by-second precision of the technique and speed of feedback are a welcome response to needs long since expressed by media producers.

The value of research during the process of media production has been acknowledged since the earliest days of educational film (as, for example, by Lashley and Watson, 1921; and by Zirbes, 1924). In 1967 it was recognised formally by Scriven under the heading 'formative evaluation'.

Figure 5(a)

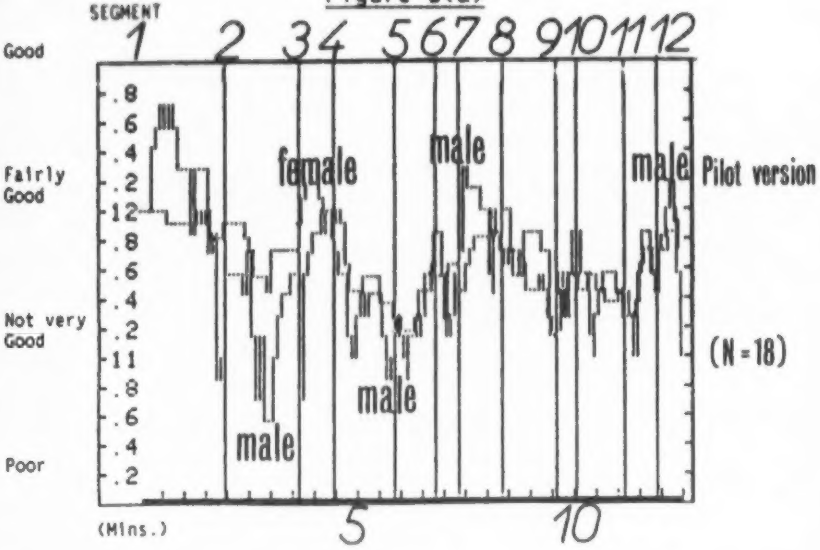
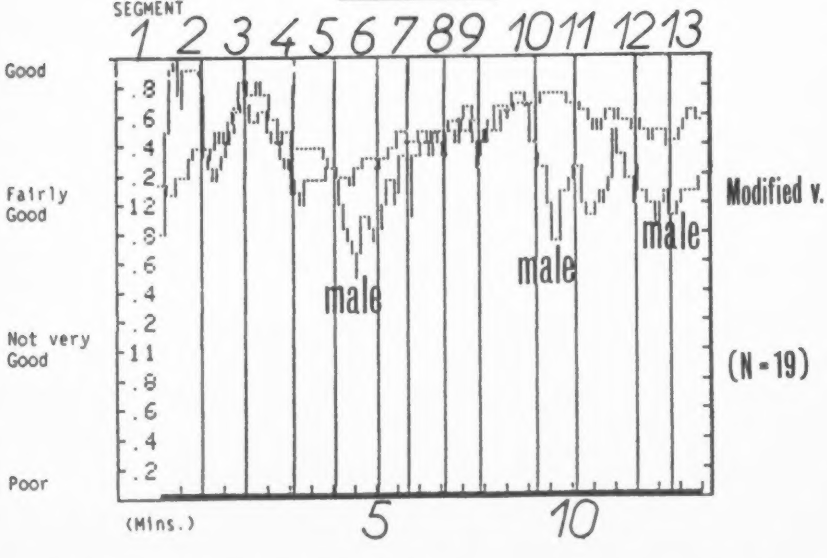


Figure 5(b)



This term has proved most valuable for the purpose of drawing a distinction between the practical types of evaluation study conducted during the production process, and the more common forms of evaluation known as 'summative', and conducted after production is completed. The latter type of study, with its tendency to expose production faults too late for producers to do anything about them, has hardly endeared the media evaluator to his production colleagues. In fact it has done much damage to their relationship.

The new response analysis techniques clearly enhance the formative evaluator's ability to work with producers on a precise and meaningful basis. Though it is certainly true that an evaluation using continual response analysis techniques may flounder for want of adequate sampling and predictability, it is equally likely to produce evidence that can be generalized to a wide range of other production situations. Even when conducted in a summative context, continual response analysis can yield evidence justifying the label 'quasi-formative' (Baggaley, 1986c). When this is the case, an evaluation study may be said to fulfil - by accident, as it were - the criteria for rigour and respectability enjoyed by traditional research studies. In previous literature, the terms 'formative evaluation and research' have been used somewhat interchangeably. Even conventional uses of the term 'formative' give rise to confusion, being used to describe the often quite distinctive types of work conducted during programme formation and concerning programme format. The label 'formative research' should perhaps be reserved for those formative and quasi-formative studies which, whether by accident or design, shed generalized light on the design of effective programme formats. 'Formative evaluation' should in turn be reserved for the less generalizable studies of individual products; they may use research methods, but they fail to satisfy accepted research criteria. On this basis, the terms 'formative evaluation and research' have quite separate uses, and the term 'formative' is unambiguous in indicating a concern with production format whether practised prior to the production process or during it.

In the history of formative research and evaluation, no technique has done more to increase the researcher's powers of inference and prediction than that of continual response analysis. In the years to come, the technique also promises to shed light on a wide range of theoretical questions, increasing our understanding of the social impact of the media, as well as helping us to design more efficient media content. Much remains to be established, however, about the design of efficient continual response measurement, and about the relationships between response behaviour and general ability. It is hoped that this chapter has helped to indicate to media producers the surprisingly specific questions that they may now ask regarding the effects of production technique, while cautioning researchers about the methodological problems still to be ironed out in the administration and interpretation of continual response studies.

REFERENCES

Baggaley, J 1980. Psychology of the TV Image. Westmead, Gower, and New York, Praeger, 190 + xv.

Baggaley, J 1985a. The analysis of media impact. B Alloway and G Mills (eds.), Aspects of Educational Technology. Vol. VIII, London, Kogan Page, 235-245.

Baggaley, J 1985b. Design of a TV character with visual appeal for pre-school children. Journal of Educational Television, Vol. 11, 41-48.

Baggaley, J 1985c. A strategy for cancer education. Report to the Canadian Cancer Society, Toronto, 2 vols.

Baggaley, J 1986a. Developing a televised health campaign: I. Smoking prevention. Media in Education and Development, Vol. 19, March, 43-47.

Baggaley, J 1986b. Developing a televised health campaign: I. Skin cancer prevention. Media in Education and Development, Vol. 19, June, 86-90.

Baggaley, J 1986c. Formative evaluation of educational television. Canadian Journal of Educational Communication, Vol. 15, 1, 29-43.

Baggaley, J and Duck, S. Dynamics of Television. Farnborough, Saxon House; and Lexington, Lexington Books, 183 + xv.

Baggaley, J and Smith, K 1982. Formative research in rural education. Media in Education and Development, Vol. 15, 173-176.

Cambre, M 1981. Historical overview of formative evaluation of instructional media products. Educational Communication and Technology Journal, Vol. 29, 3-25.

Carpenter, C, Eggleton, R, John, F and Cannon, J 1950. The Film Analyser. Port Washington, New York, U.S. Naval Training Services Center.

Cirlin, B and Peterman, J 1947. Pre-testing a motion picture: a case history. Journal of Social Issues, Vol. 3, 39-41.

Clarke, A and Ellgring, H 1983. Computer-aided video. P Dowrick and S Biggs (eds.), Using Video. New York, Wiley, 47-60.

Conger, R and McLeod, D 1977. Social interaction in small groups. Behaviour Research Methods and Instrumentation, Vol. 9, 418-424.

Corporation for Public Broadcasting 1981. A Comparison of Three Research Methodologies for Pilot Testing New Television Programs. Office of Communication Research, CPB, 132.

Lazarsfeld, P and Stanton, F (eds.) 1944. Radio Research, 1942-43. New York, Duell, Sloan and Pearce.

Malik, M 1984. Biometric communication research for television. J Baggaley (ed.), Experimental Research in TV Instruction, Vol. 4, Montreal, Concordia University, 85-105.

Nickerson, R 1979. Formative evaluation of instructional TV programming using the Program Evaluation Analysis Computer (PEAC). J Baggaley (ed.), Experimental Research in TV Instruction. Vol. 2, St. Johns, Memorial University of Newfoundland, 121-125.

Sawin, D, Langlois, J and Leitner, E 1977. Parent-infant interactions. Behaviour Research Methods and Instrumentation, Vol. 9, 425-428.

Scott, K and Masi, W 1977. Infant visual behaviour. Behaviour Research Methods and Instrumentation, Vol. 9, 429-433.

Scriven, M 1967. The methodology of evaluation. B Worthen and J Sanders (eds.), Educational Evaluation: Theory and Practice. Worthington, Ohio, Charles A Jones.

Sykes, R 1977. Member interaction in a large spontaneous group. Behaviour Research Methods and Instrumentation, Vol. 9, 429-433.

Toronto Globe and Mail 1985. TV device enhances market research, 22nd May.

Lashley, K and Watson, J 1921. A psychological study of motion pictures in relation to venereal disease campaigns. Social Hygiene, Vol. 7, 181-219.

Wingerson, L 1982. Nice guys finish first. Discover, March, 66-67.

Woodworth, R and Schlosberg, H 1961. Experimental Psychology, (3rd edition revised). London, Methuen, 948 + xi.

Zirbes, L 1924. Relation of visual aids to educational objectives. National Education Association Addresses and Proceedings, Vol. 62, 964-966.

FORMATIVE EVALUATION OF ETV IN THE EARLY STAGES OF PROGRAMME DEVELOPMENT

A Duby

Introduction

Evaluation experts, as early as in the 1960's, called attention to the need to distinguish between formative and summative evaluation within the framework of programme development (Cronbach, 1963; Scriven, 1967). This distinction is almost universally accepted: most empirical studies are labelled either formative or summative. However, the experience accumulated with the evaluation of educational television at the SABC justifies a further categorization of formative evaluation activities. Formative evaluation is carried out during the whole process of programme development; this process consists of several clearly distinguishable phases. As the development activities vary from one phase to another, it follows that decision-makers face different questions at different phases. Consequently, the task of the evaluator differs according to the development phases of the programme.

The formative evaluation process starts when the programme is conceived, and stops when full-scale production begins. Figure 1 provides an overview of six distinguishable phases of programme development. An attempt will be made to explain the unique characteristics of these six phases of development, and also to suggest what kind of evaluation activities could be, and should be, carried out in conjunction with these development activities.

In the figure, the formative evaluation of an educational TV programme (ETVP) is divided into two stages a) formative evaluation (F.E.) in the early stages of ETVP development and b) F.E. conducted in the later stages of ETVP development.

FIGURE 1: INTEGRATING FORMATIVE EVALUATION WITH ETV DEVELOPMENT

FORMATIVE EVALUATION	PHASES	ETV STAGE OF DEVELOPMENT
a. Early F.E.	I Preparation	1. Idea 2. Rationale
	II Development	3. Content 4. Script
b. Late F.E.	III Production	5. Production of pilots 6. Full-scale production

This chapter is concerned mainly with early formative evaluation incorporated in the development stages of the ETVP process (phase II). It is also concerned with self-formative evaluation (S.F.E.)(Duby, 1985b), which is conducted by both ETV developers and evaluators in the process of improving ETV-programmes, and to assist TV professional staff in self-development.

Methodological Features of Formative Evaluation

Empirical research on the usefulness of formative evaluation studies was carried out by Abedor (1972). He defined formative programme evaluation as an attempt to answer three questions: how to identify weak points in a programme through data collection; how to analyze these data to revise hypotheses; and how to design and evaluate the revisions.

Although Scriven (1967) claims there is no difference between formative and summative evaluation methodologies, some methodological features are used in formative evaluation which are not used traditionally in summative evaluation studies.

Markele (1967) observed the behaviour and reactions of a single learner and drew conclusions about the quality of the programme used, instead of testing a mass of subjects to obtain statistical data. He also demonstrated the merits of trying out materials at the draft stage with a single, or a few learners, using a non-control-group design. Cronbach (1963) emphasised the need to analyze the draft of an instructional programme to revise it after a trial, and to provide developers with evaluation evidence during the developmental process. He gives the following reason: "evaluation used to improve the course while it it still fluid contributes more to improvement of education than evaluation used to appraise a product already placed on the market" (p 675). He also proposed the examination of the effectiveness of segments of a programme rather than the "course as a whole" in order to identify "places where the course can be strengthened".

Alkin (1974) focused on the testing of individual components of a programme as a typical formative evaluation activity. Light and Reynolds (1972) suggested a mëthod by which the evaluator can discover weaknesses to effect improvements in children's programmes.

According to some researchers (Mann, 1969; Werner, 1980), utilization of expert judgement is conspicuously missing in evaluation reports. Yet, the formative evaluation of "Sesame Street" and "3-2-1-contact" ETVPs relied heavily on expert opinion (Cook and Crutin, 1985; Palmer, 1978). Cook and Curtin (1985) claim that the adaptation of the Children's Television Workshop (CTW) model by ES-TV was very poor partly because expertise was not used.

Cook et al (1975), in dealing with the experimental formative evaluation of the "Sesame Street" programmes, indicated that these evaluation activities were particularly beneficial as the evaluators could not rely on any well-established theory in the process of programme development. They also characterized formative evaluation as an activity typically concerned with a low level of generalizability: in their view generalizability is not as important in formative evaluation as is the speed of feedback.

Alkin (1974) supplied additional features for doing formative evaluation studies: their design should be explanatory and flexible; they should emphasise interactive processes, they should make use of a variety of research instruments, but not make use of comparative studies.

Needs Assessment

The first source of criteria to be considered by the educational programme developer is the assessed needs of the target audience (Suarez, 1981); consequently, assessing needs can be considered as a component part of formative evaluation.

Scriven (1974) has defined needs assessment as "good evidence that the product fills a genuine need or - a weaker alternative - a defensible want: true needs assessment involves establishing that the product actually facilitates survival, health, or some other defensible end that is not now adequately serviced (p 9). Scriven suggests a five-point scale for prioritizing needs: "Maximum priority (4), great importance (3), probably significant need (2), possible significant need (1), and no real evidence of need (O)"(p 9).

In identifying and prioritizing audience needs one is concerned with an analysis of audience contentment, and the selection and justification of the use of TV as a suitable medium for conveying the message to fulfil the needs. Needs assessment research, as a preparatory study proceeding evaluation activities, has also been described by Provus (1977), Palmer (1974, 1978) and Cook and Curtin (1985). Assessing needs is an attempt to find out what the target audience must understand in order to master the content of a programme; what needs have already been met; and the discrepancies between the desired outcomes and the actual conditions.

Try-Out Studies

Another type of formative evaluation is the try-out study which involves trying out a short programme, or segments of programmes, with a sample of the target audience. This process is widely used by ETV production companies and is described by many researchers, e.g. Baggaley and Duck (1976), Land (1972), Lesser (1974), Mates (1980), Nickerson (1979), Palmer (1978), Tidhar (1979), etc. Although try-out studies brought the CTW its success, some production houses have failed to implement the CTW model. They have focused on pilot-programmes try-out studies rather than on the continuation and iteration of formative evaluation activities during the development stages of a programme (Cook and Curtin, 1985).

Cook and Curtin (1985) gave the following five reasons for the success of the CTW model:

- The funding of long pre-broadcast periods in which "need analysis is conducted, objectives are made explicit, initial plans are critiqued, and segments and pilots are produced, tested and modified" (p 95).
- The establishment of co-operative teams in which producers pay close attention to formative evaluators and subject experts, and use their advice to make programme decisions.
- Producers continue to pay attention to formative evaluation feedback during the full production phase.
- Formative evaluation researchers and content experts are useful, for without them the pedagogic or artistic quality of productions would be lower.
- High quality of pilot programmes produced after formative evaluation is guaranteed, and they are accepted by the target audience.

In their summary review, they claim that "no feature of 'Sesame Street' was adopted with more enthusiasm than the part dealing with how continuous formative evaluation, substantive knowledge skills can interact to increase the media production quality" (p 86). "The model was adopted so that it is now virtually impossible to find any educational television without a formative research component of some sort or another" (p 99).

Two innovative features of the formative evaluation activities used in the CTW model can be emphasized here: iterative formative evaluation feedback and continuity of in-house staff involvement in the formative evaluation/development process. These two features contributed to the instant and constant improvement of the product and the staff involved.

Self-Formative Evaluation

Self-formative evaluation of instructional media takes place during the process of programme development by the users or developers of the programme or by both jointly (Duby, 1985b). It is not to be confused with the type of self-evaluation which addresses itself to psycho-therapeutic self-analysis. It is also different from evaluation conducted by an outside researcher. Although the self-formative evaluation concept is hardly touched on in the literature on formative evaluation (Nevo, 1983) or in the evaluation of TV, in some surveys the issue of evaluation done by internal evaluators is discussed.

Abedor (1972) has shown that the quality of new materials is gradually improved after the developer's work is submitted for formative evaluation. After a learning period during which three or four units are scrutinized and revised, the developer reaches a level at which only small revisions to her drafts are needed.

Lewy (1985b) scrutinized the gains from in-house and external evaluations and concluded that formative evaluation can be done best by internal evaluators. The chances of close co-operation and simultaneous planning of development and evaluation activities by developers and evaluators are the main common features of in-house evaluation (Lewy, 1985b) and self-formative evaluation (Duby, 1985b). The use of self-evaluators or staff evaluators in the formative evaluation of ETV has yielded the following outcomes (summarized by Duby, 1985b):

- The threat associated with the activities of an external evaluator is reduced.
- An individual who has carried out self-evaluation activities tends to be more accepting of other evaluations.
- Self-evaluation increases awareness of the advantages of evaluations.

- The party carrying out the self-evaluation has first-hand knowledge of the actual evaluation needs and is familiar with the context. (This may effect the validity of the findings).
- Self-evaluation carried out jointly by developers and users strengthens communication between them.
- Self-evaluation is a powerful tool for self-improvement.
- The producer who also functions as an evaluator is constantly informed of the evaluation results, and is able to make decisions for the improvement of the programme.
- The user himself is involved in the evaluation and is therefore better motivated to play an active role in producing educational TV programmes.
- Information obtained from formative evaluation is assimilated into the production process quickly.
- The involvement of the ETV-developer and the ETVP-user in the production process ensures the maximum suitability of the programme for its intended functions.
- An ETVP which is evaluated in the early stages will cost less, and will be produced faster than an ETVP which is evaluated in the later stages of production.

Formative Evaluation Activities at Each Stage of ETVP Development

EARLY FORMATIVE EVALUATION

As indicated in Figure 1, the early formative evaluation activities concern the first four development stages of ETV.

When the ETVP is only at the concept stage, one can examine whether there is a need for such a programme. At this stage the manager in charge of programmes can employ researchers to provide him with information on whether other teaching aids have been used to present the subject, whether the programme focuses on the needs of a defined target group, and possibly suggest a delivery format (broadcasting, video cassettes, teletex, etc). The CTW researchers also did a need analysis to tease out detailed behavioural objectives linked to the TVP funder's educational objectives (Cook and Curtin, 1985).

Cost Analysis

This is also the time to conduct an analysis of cost-effectiveness. The purpose of this type of analysis is to determine what a programme costs and how effective it is. Effectiveness is often assessed in terms of rewards.

The producer does not provide an automatic answer to the question: Is the programme worth its costs? The evaluator will have to assess the needs in terms of the costs to obtain an answer which may not be unequivocal, but which clarifies the choices considerably (Scriven, 1981).

Availability of Data

At this stage, subject experts can narrow the scope of the target audience, and provide data regarding its needs and abilities.

The needs assessment, cost analysis and expert advice are used to write the rationale of the programme (Figure 1, Stage 2). The nature of the information gathered will enable, at this stage, a decision to be taken on the production of non-effective programmes.

The content is first decided on by the manager and the programme writer, based upon the rationale (Figure 1, Stage 3). Subject experts are invited to provide information on content and, in order of importance, the educational objectives.

Subject experts will continually provide the producers with advice on the scientific accuracy of the content and its acceptability to the target audience.

At the second (rationale) stage, only data external to the programme are used. Internal programme information is generated in the third (content) and the fourth (script) stages. Evaluators then try out the content and adapt it in line with their evaluations. The data used for the evaluation will comprise an assessment of the script content, a comparison of the documented script with the rationale, and with the content, to check whether scientific and behavioural objectives have been incorporated. Although the medium (TV) is still not ready for transmission, production and target-audience issues can, to some extent, be predicted. The criteria to be predicted are: suitability of the programme format to content, potential of the medium to fulfil the subject requirements, and suitability of a programme to target audience needs. The ability of the target audience to accept, absorb and use the ETV message can thus be used as an evaluation yardstick in the early stages. Another external source of information for assessment is the published literature.

The rationale, the documented content of the programme, the written script, the written internal reports (based on expert input and the defined objectives) are a collection of internal information sources available to the evaluator in the early stages of the ETV development. External sources of data can be established at this stage by enlarging the self-evaluation team. Duby (1985b, and c), showed that it is possible at this stage to collect valid and credible data from potential users of the programme (e.g. teachers and TV professionals) on comprehension and acceptability.

Duby (1985b) found that teachers (potential users of ETVP, n=90) and TV professionals (TV managers, researchers and producers, n=12) were not only willing to be involved in formative evaluation activities, but were also able to supply the evaluator with worthwhile data.

The aspects examined with these subjects were: coverage of course objectives, suitability of programme to viewer needs and abilities, and suitability regarding programmes for pre-school and secondary school pupils. It was found that teachers could provide information on the analysis of the content of programmes given to them in a script format. The questions administered to the TV professionals dealt with coverage of course objectives, fit with classroom setting, and language difficulties. The questionnaire was given to them twice, firstly for the evaluation of the ETVP script, and then after viewing the produced ETVP.

It was found that the TV professionals did not change their script-based evaluation after viewing the programme (98 percent consistency): "judgement issued at this stage is mostly sustained at later stages of the programme's development" (Duby, 1985b:169). The researchers could not make a valid judgement of ETV features at the 'script stage' (before viewing the programme) and, thus, this procedure is not suggested as a source of data at the 'script' stage.

Content Analysis

Content analysis is defined by Scriven (1981) as "the evaluative or pre-evaluative process of systematically determining the characteristics of a body of materials." (P. 31).

The techniques for this evaluation process are mainly: counting frequency of certain kinds of words; analysing the plot in order to determine balanced appearance of certain images, e.g. males, females, children, robots, blacks or whites etc; monitoring reinforced main ideas or main objectives against side issues, etc.

Its main use is to ensure that the right learning effects (pay-offs) accrue as a result of exposure to the right amount of selected effects. Thus content analysis can result in content validity which indicates whether the content appears to meet all the requirements for congruence between claimed and actual content, only claimed requirements and/or meets unclaimed objectives.

This semi-quantitive, semi-qualitive examination technique is applicable to our study as it can be applied at the 'content' level, mainly at the 'script' stage and also when the pilot programme is available.

LATE FORMATIVE EVALUATION

The information sources mentioned above are insufficient when production issues such as format, presentation, sound, pace, etc. have to be tested, which are best examined by viewing a video recording (Baggaley and Duck, 1976; van Vuuren, 1985; Cook and Curtin, 1985; Cook et al, 1975). For this reason it is common for ETV production houses to produce segments, pilot shows or pilot programmes for pre-viewing and formative evaluation (Baggaley et al, 1976; Cook and Curtin, 1985; Cook et al, 1975; Tidhar, 1979). These formative evaluation activities are labled here 'late formative evaluation'. In the CTW, for example, preliminary programme segments produced after the script had been written were evaluated by trying out the segments of the programme with a target audience. The segments were then modified and one or more pilot programmes (completed episodes), were produced and tried out on viewers (Cook and Curtin, 1985). This was the first opportunity to test production issues in a real-life situation. An example of a late formative evaluation exercise to emphasise this follows.

A pilot programme for the series called "Ziki Zikombot" to teach $5^1/_2$ - 7 years olds basic mathematical concepts was produced by the SABC. The producers of this programme thought that the series may be appropriate to capture other viewers, and increase the size of their audience. The following design (Duby, 1985c) was used to test the pilot programme with a variety of potential viewers.

The objective of the study was to examine some issues of production in order to ensure that the salient points of the programme were not lost on the group. The hypothesis that "the more attractive, comprehensible and acceptable the pilot programme, the greater the potential effectiveness of the series" (Rydin, 1973) was adopted.

Table 1: Ages of Subjects

Group	Ages of Subjects Years	Number of Subjects (n)
I	3 - 4	12
II	4 - 5	20
III	5 - 6	26
IV	6 - 7	19

The data revealed that all the children were attracted by the 'robot' and by the 'child-actors'. The children gave these aspects full attention whenever they appeared on the screen and expressed affection for them. Their attention was drawn to the sound made by the robot, and it did not irritate them. (Although the sound was accepted by the children, further tests should be conducted to ascertain whether they could actually hear the words).

The children's attitude to the adult who acted as the 'professor' was not the same across the board. Those aged 3-6 were not attracted to the professor whereas the oldest group (6-7) was. The other two adult actors did not attract significantly more attention, although no hostile reaction was recorded. The 'space set' was very appealing to the young children (3-6) but unimportant to the older group (6-7). The electronic gimmicks were very appealing to the very young children (3-5) and fairly appealing to the others.

The oldest group (6-7) showed great interest in watching the 'real' material (EBE material), which was not appealing to the other groups. Although the group aged 6-7 did not pay much attention to the logo, the three younger groups were very interested and carried on watching the screen for as long as the logo was shown. The programme was appealing and entertaining to children aged 3-7. The sets, acting, sound and visual effects were very well accepted by the 3-5 year olds.

The adult actor, scientific visuals and verbal information were well accepted by the children aged 5-7.

By narrowing the target audience to one age group, one can make the teaching and entertainment aspects of the programme more effective (Duby, 1985c). The point to be emphasised in this chapter is that such issues as the attractiveness of the programme could not be tested before the programme was actually produced. Editing techniques are also the subject of later formative evaluation, for example, concepts like 'round' and the statement that the sun is made of fire were emphasised in the ETVP 'Ziki Zikombot' through verbal repetition and visual reinforcement, and were recalled by the children after viewing. 'What the moon is made of' was not repeated in the programme and could not be recalled by many after viewing. By intentionally identifying concepts, and deliberately repeating them, using visuals and sound, one can reinforce the comprehension and the memorization of educational elements (Rydim, 1973). Editing issues such as 'bulletin repetition' should be looked for only when a pilot programme is being developed.

The Ziki Zikombot formative evaluation outcome of the programme were appealing and acceptable for all the characters, sets and content. However, the main recommendations for improvement concern comprehension. This aspect should, and could have been dealt with when the content and script were available. If formative evaluation had been done in the early stages, the script could have been rectified easily. The issues could not be rectified late and were left uncorrected when full production started. It will be established later in this chapter that early formative evaluation is also a way to save time and money in the long run as early evaluation can obviate unnecesssary reproductions.

There is no doubt that some issues cannot be evaluated in the early stages of TV production because "as one proceeds with the development of the programme, more data becomes available for evaluation" (Duby, 1985b), moreover there are some issues which cannot be tested or even predicted before a video programme is produced.

Nevertheless, the main point of this chapter is to show that a great amount of valid and credible data can be gathered at the preparation stage of ETVP development (idea and rationale) and mainly at the development pre-production stage (content and script), before real production starts. The testing of a pilot programme can thus be limited to the testing of production features and the acceptance of the programme by the target audience.

Justification for Early Formative Evaluation

Because many programme participants are involved, formative evaluation activities at all stages is time-consuming, but relatively inexpensive. It has been calculated that a maximum of 30 percent of the production budget is spent on the first two stages. This figure includes experts' fees, script-writing, correction of contents and correction of script. Once the real production begins, the amount of money spent per stage increases and is used for the set, the shooting of the recording and tape editing (70 percent of the budget). At the production stage, time becomes a factor as well, because the equipment, the set and TV professionals arê expensive resources. Time and financial expenses in ETV production (based on the ERTVI, SABC 1984 budget report, Duby, 1985a) are summarised in Table 2.

Table 2: Time and Cost in TV Processing

STAGE OF ETV	TIME	COSTS
Preparation and Development	70%	30%
Production	30%	70%

(Duby, 1985a)

Changing course materials on videotape is more expensive than changing them on paper. This is one of the advantages of early formative evaluation. The amount of data collected for evaluation is small in the early stages: it increases substantially when production begins (This is probably the reason why evaluation of pilot programmes is so popular).

Two more major factors must be taken into account in early formative evaluations: Whose evaluation input will produce the most valuable evaluation input at each stage? Which criteria will produce the most meaningful input at each stage?

The Value of Evaluation Inputs at Each Stage

There is a great deal of discussion on whom should be involved in the process of evaluation. The CTW evaluation model is presented in Figure 2 (Cook and Curtin, 1985). The titles of some participants in the evaluation process are in parenthesis to indicate that, although they were involved in formative evaluation, they also exhibited their willingness to work with researchers, and were ready to implement evaluation recommendations. Yet the extent of their involvement is not shown.

FIGURE 2: FORMATIVE EVALUATION - THE CTW MODEL

FORMATIVE EVALUATION ACTIVITIES	PARTICIPANTS IN THE EVALUATION PROCESS
1. *Pre-Production Phase*	Experts, internal & external researchers
1.1 *Experts* conduct *needs analysis* and hammer out detailed *behavioural objectives linked* to the *founder's definition* of the educational problem to be tackled.	
1.2 Behavioural objectives are agreed upon after many iterations.	Internal executive producer, experts
1.3 Producers and other production staff are hired for their proven creative skills and their *willingness to work* with *researchers*.	
1.4 Developing production philosophy or format.	Researchers and Experts
2. *Preliminary Production*	Internal researchers
2.1 Formative evaluation (laboratory style) of comprehensibility and appeal of *segments* to viewer population and modification of segments.	
2.2 Pilot show (one or more) to estimate audience appeal: comprehensibility to be tested in the field. (Interview, questionnaire, mechanical). To select best pilot programme and best features.	
2.3 Discussing the findings: Informal, supportive, teamwork.	Researcher(Executive producer, writers) and content experts.
3. *Production*	Production staff, internal researchers & subject experts
3.1 Full scale production which receives permanent feedback from researchers and experts.	

The argument in this chapter is that the more in-house production staff are involved in the evaluation process - in hammering out weak points, in initiating evaluation activities and in analysing problem situations - the more open they are to recommendations from evaluation studies and the more likely the studies are to improve their professional expertise and understanding, and lead to the application of the evaluation results (Duby, 1985b). Examples of the involvement of in-house staff in evaluation are presented below:

IN-HOUSE STAFF

In 1985, Duby conducted a feasibility study at the SABC to determine the ability and willingness of in-house staff to do self-formative evaluation. The subjects included three types of specialists: TV managers, producers and internal researchers. They completed a questionnaire on an ETV script, before and after viewing a programme. Approximately 85 percent of the respondents considered it appropriate that changes and corrections were made during the production process. Table 3 shows their opinions on when evaluation should take place.

The TV professionals were of the opinion that

- formative evaluation should take place at different stages of production rather than after production;
- formative evaluation should be done mainly during the writing and the script stages;
- evaluation can also be done soon after the ideas and concepts of the programme have been defined.

One may develop an index for measuring the inclination of groups of TV professionals to use formative evaluation.

The responses in Table 3 can be represented as an index of number of responses in favour of formative evaluation to total number of possible responses. The index is shown in Table 4.

Table 3: Responses to questions indicating at which stages formative evaluation should take place

	NUMBER OF PERSONS			
	MANAGERS (n = 3)	PRODUCERS (n = 4)	RESEARCHERS (n = 5)	TOTAL (n = 12)
Idea	3	1	5	9
Rationale	2	3	5	10
Content	3	4	5	12
Script	3	4	5	12
Pilot Programme	3	2	5	10

Table 4: Proportion of respondents interviewed in favour of formative evaluation of ETVP

	TYPE OF EVALUATOR		
	MANAGERS (n = 3)	PRODUCERS (n = 4)	RESEARCHERS (n = 5)
Maximum (persons)	15	20	25
Obtained (persons)	14	14	25
Obtained (percent of maximum)	93%	70%	100%

Producers were less inclined to use formative evaluation than were TV managers or researchers. Producers were also not keen to use any other form of evaluation.

In Germany, Sweden, Israel and South Africa it has been found that producers generally have negative attitudes .to research. The CTW hired producers who were willing and able to work with researchers.

Table 5 gives the TV professionals' opinions on who should be involved in the evaluation process.

Table 5 shows that

- each particular group of professionals tended to emphasise its own role in the process of evaluation;
- managers were willing to co-operate with internal researchers, internal evaluators, external evaluators and subject-matter experts. The producers were also keen to co-operate with external experts.
- researchers did not regard other expert groups as indispensable contributors to the evaluation.
- managers were willing to involve all the professionals in the evaluation of their programme.

The in-house SABC RTV1 production team's attitude towards formative evaluation in the early stages of production was positive.

The results reported here were obtained after long in-house overt and covert in-service training activities to persuade the ETV staff that formative evaluation is a desirable activity. The author has also trained teachers, users and evaluators of instructional materials in evaluation skills. It appears that training sessions of 2-3 days in duration can make teachers less resistant to formative evaluation, increase their openness to evaluation, and improve their evaluation skills (Duby, 1985b).

Table 5: Responses to the question regarding the persons who should be involved in evaluation

PERSONS TO BE INVOLVED	RESPONDENTS			
	MANAGERS (n = 3)	PRODUCERS (n = 4)	RESEARCHERS (n = 5)	TOTAL (n = 12)
1. Manager	3	2	1	6
2. Producer	2	4	4	10
3. Internal Researcher	3	4	5	12
4. External Evaluator	3	0	0	3
5. Subject Matter Expert	3	3	1	7
6. Educator (Expert)	2	3	1	6

Table 5: Responses to the question regarding the persons who should be involved in evaluation (continue)

7. Psychologist (Expert)	1	0	1	2
8. Ad hoc Comm of Experts	2	1	4	7
9. Educational Practitioner	2	1	4	7
10 Target Audience Reps.	2	2	4	8
TOTAL	26	21	26	73
Maximum = nx10	30	40	50	120

(Duby, 1985b)

In the same study TV professionals' ability to do self-formative evaluation of TV programmes was examined.

They made meaningful judgements of the script, and to some extent also of the medium, but one can hardly "attribute to the TV professionals a competency to judge whether a particular programme fits the developmental level of the target viewer" (Duby, 1985b:175). They were also "reluctant to acknowledge that familiarity with a concept derived from educational psychology is an indispensable prerequisite for producing satisfactory programmes" (Duby, 1985b:183), and that they "should receive more intensive training to acquire a level of mastery of this important ability" (Duby, 1985b:177). The TV professionals showed a lack of self-confidence in dealing with learners and did not recognize the importance of concepts derived from educational psychology. Nevertheless, by being involved in empirical studies, the producer was in direct contact with her target audience. She became more research orientated and more open to criticism and self-improvement.

Because of their knowledge of the content of the script, the in-house production team should be involved in, and consulted on, the self-formative evaluation process.

TEACHERS

Teachers, with their knowledge of learners' needs, capabilities and attitudes towards learning, should become active and selective users of ETVP. Duby (1985b) found that teachers can make a contribution to evaluating the script. "Teachers felt comfortably able to make statements about the educational aspects of the content and had a high level of self-confidence in using terms derived from educational psychology, such as the pre-requisite and the taxonomic categories of educational objectives.

They revealed competence in evaluating the fit between learner characteristics and programme content but less so in making evaluation comments on media characteristics and on their match with the content (Duby, 1985b).

As teachers, managers, researchers and producers are willing to take part in the formative evaluation and can make valuable inputs, short workshops should be held with them to focus on the cognitive, affective and psychomotoric aspects of the content, the learner and the medium, and their effects on each stage of ETVP development.

EXPERTS

As indicated before, managers, producers and researchers at the SABC ERTV1 had different ideas on who should be involved in the process of evaluation. Nevertheless, the managers revealed the highest level of openness to the involvement of experts of various types in the evaluation.

Table 5 shows that the producers were resistant to external consultants, whereas internal researchers were inclined to use subject-matter experts only, and rather democratically i.e. brainstorming sessions and ad hoc committee meetings. One can understand this complex attitude in the light of the variety of experts participating in the production of TVP. In addition, timing, financial, social and psychological factors are combined to create in the production house one of the most pressured processes known. It is thus suggested that the number of experts be limited by sampling.

All experts participating in the production of educational materials should have some knowledge of education principles. Acquisition of such knowledge will make them more understanding of the product they prepare: it will enable them to communicate better with other experts participating in the production of ETV and with ETV users (Fox, 1985).

"Nevertheless, experts of various types should not tresspass on the boundaries of expertise of other types and should not make decisions related to a domain of speciality which lies outside their domain of expertise" (Duby, 1985b:186). Consequently the internal evaluator should take a leadership role and should compile the evaluation report, discuss the results with the producers and managers, and should motivate them to accept the report and interpret it. The evaluator should also be responsible for ascertaining the validity and the reliability of the study. This will "obviate the waste of resources used for implementing untested strategies attempting to achieve unwarranted goals" (Nevo, 1985:3).

An eductional TV programme is a complicated project in which the medium as well as the learner's needs, attitudes and abilities are considered, and the content accurately and appropriately portrayed. Psychologists, educators, sociologists, experts in the relevant subject matter, technical experts and experts in relevant inter-disciplinary fields should be called in to advise during the process of production. (Cronbach, 1963; Leide, 1977; Lewy, 1973).

Nevo (1985) found that synthesized experts' opinions were perceived by the evaluation clients as very relevant information. The clients appreciated the relatively speed with which they received the first evaluation feedback, which "served as a good start for the whole evaluation project" (p 12). Nevo suggests that experts with extreme views for and against the project be precluded from giving advice. However, persons who take an extreme position may also be the most knowledgeable in the subject; moreover, the experts involved in the project may be the only people available in a particular subject. Sometimes it is better for the instructor to choose her own experts to help her convey a new, important (scientifically supported) educational message. Experts should not only have formal qualifications but should know their subject discipline.

A team of experts could comprise academicians and practitioners in the field, drawn for the subject discipline as well as from interdisciplinary subjects. For example, for a TV programme on teaching elementary geometrical concepts to pre-school children, those who should be invited to contribute include an expert in the field of mathematics, a psychologist specialising in early childhood and a nursery-school teacher who can advise on the pace, level of difficulty, attractiveness, etc. of the programme. Nevo (1985) has recommended that the number of experts assigned to any one project be limited to 2-10, depending on the complexity of the project, its size and the resources available.

Duby (1985a) distinguished between two functional kinds of expert input: 'judgemental statements' similar to those described by Stake (1967) and Fox (1971) - their judgemental statements should stem from a comparison of the evaluation material with standards (Lewy, 1971, 1973) - and 'expert opinion' which stems from the experts' established and recognized knowledge and experience. As the use of expert input serves, inter alia, to compensate for the lack of information available in the early stages of the development of the project, the above differentiation can help the evaluator. An example of a question the evaluator can ask is: In your opinion, will pre-school children understand the message conveyed by the robot in the included episode, or is it correct that the sun is round and is made of fire? (Duby, 1985c).

In the following figure an attempt will be made to classify experts' contributions according to their type of advice.

The information contained in Figure 3 indicates that, for most of the issues, the experts' input is based on judgements and not on opinions. Thus one can ignore the fact that experts have been known for their subjectivity and conclusiveness (Scriven, 1981), and should include experts' judgement as a powerful source of information for formative evaluation.

FIGURE 3: TYPES OF EXPERT CONTRIBUTIONS

ISSUE TO BE EVALUATED BY EXPERTS (Duby 1985b, Lewy 1973, Nevo 1985)	JUDGEMENT	OPINION
* Significance of topics		+
* Fit of planned activities to programme specifications	+	
* Fair representation of different views	+	+
* Scientific accuracy	+	
* Correctness of language	+	
* Can an opportunity to learn be found in the programme?	+	
* Relevant information on existing values	+	
* Information on standards relevant to worth and merit of the evaluated object	+	
* Information responsive to concerns of relevant audiences	+	
* Will the programme be used properly? (Prediction)		+
* Has the programme an appropriate level of difficulty?		+

Other drawbacks to using experts, stem from the fact that, in the early stages of programme development, it is difficult for the evaluator to present the project to the experts and to define questions for the experts. The objectives of the programme are still emerging in the early stages, and thus it is difficult to finalise and define. In addition, sometimes a suitable expert in a certain field cannot be found, or her co-operation cannot be obtained (Nevo, 1985). Despite these drawbacks, and since experts' advice is the only evaluation tool available in the early stages, the employment of experts in these stages is highly recommended. The advantages of using expert input as an evaluation tool in the early development stage of a programme can be summarised as follows:

* The use of the experts' contribution serves to broaden the scope of information collected on each evaluation programme (Nevo, 1985).

* By using expert opinion in the production stage, one gains an extra tool for evaluation which could help to improve the programme (Duby, 1985a).

* In an educational programme we cannot afford to expose our audience to mistakes; experts can prevent mistakes (Duby, 1985a).

* It is cheaper in terms of time and labour to employ experts than it is to try out a pilot or a final programme.

* In the early stage of programme development, expert evaluation can help to detect problems (Lewy, 1973; Nevo, 1985).

* In the very early stages of production, there is few other usable methods of evaluation.

* In the early stages of programme development, producers are still open-minded to criticism, whereas later they tend to want to justify their products (Lewy, 1973).

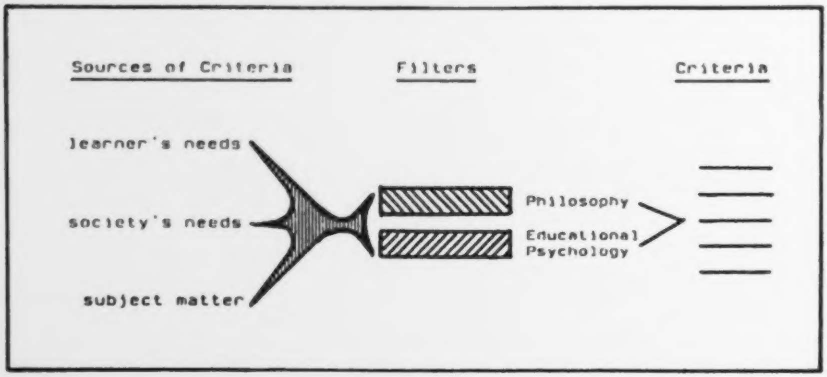
* It is easier to manipulate the time schedule of experts than of an audience for try-outs, observations and tests (Lewy, 1973).

Criteria for ETV Evaluation

Scriven (1974) follows Stephens, a famous yacht designer, who published a '100 - item checklist to be used in the evaluation of racing and cruising yachts' and suggests that for each educational project a 'Production Evaluation Checklist' (p 7) should be compiled. "The status of every item in that checklist can be expressed by saying that each is a 'desiratum' and some items are 'necessitata' ... Each of these conditions must be met in order that one should have impeccable ground for a conclusion of merit for an educational programme" (p 8).

Formalization of educational objectives does not occur in a vacuum, disconnected from reality, but as a result of socio-cultural-economic factors, and should reflect the values and principles of a given society (Postlethwaite, 1977). What is to be measured when educational objectives are determined? The systematic approach to this question was discussed firstly by Tyler in 1940, and it is still one of the fundamental questions. Tyler (1950) mentions three sources for determining objectives - the society, the learner's needs and the subject matter - and two filters for the final selection of objectives: philosophy and educational psychology. Tyler's model can be described schematically as follows:

FIGURE 4: SOURCES OF CRITERIA ACCORDING TO TYLER'S MODEL



Many researchers oppose Tyler's approach. Eisner (1974) for example gives four reasons for not specifying objectives for curriculum planning:

- The teaching process is dynamic and complicated and results in a variety of outputs which are difficult to predict and therefore impossible to prescribe as objectives.
- In certain areas, such as art or creative writing, the learner is asked to show initiative, positive attitudes, originality, and innovation, all of which are very difficult to prescribe.
- Not all educational products are measurable.
- A desired action or pre-determined content may be the basis for an educational programme and not necessarily a final determined objective.

In stating this, Eisner (1974) formulated the fundamentals for those who believed in the freedom of the educationalist to act spontaneously and not be constrained by predicated educational objectives.

Many researchers have tried to bridge the gap between the two approaches. Domingo (1979) claimed that it is important for educational objectives to be predicted for devising a curriculum, but one must be aware that determined objectives do not contain all that has been achieved by the learner following a specific curriculum.

In their article, 'Issues in the Evaluation of Educational Television programmes', Aversa and Froman (1978) describe an evaluation model. This model is based on three categories of criteria: content, design and aesthetic quality which are found among the three major persons on ETV teams: the content expert, the instructional designer and the producer. The ETV criteria are displayed in the following exhibit.

FIGURE 5: CRITERIA FOR EVALUATION OF TELEVISION PROGRAMMES

Content Criteria

There are content errors.
There are ambiguous content areas.
The content is important.
There is an appropriate amount of content.
The level of the programme is appropriate to the level of the course.
The programme presents interesting and challenging ideas.

Design Criteria

The programme achieves its intended purpose.
The programme is either too simplistic or too difficult.
The programme is well organised (overview, transitions, summaries).
The information is clearly presented in the programme.
There are too many or too few topics.
The programme develops the main points adequately.
The main points are well illustrated.
Non-examples of the main points are used.

Aesthetic Quality (Production) Criteria

There is a good balance between in-studio and on-location filming.
The voice-over technique is used effectively.
The programme is interesting.
Existing stills and footage are put to good use.
There is continuity between sequences.
The programme is paced too fast or too slow.
The sequences are of an appropriate length.
The level of comment made by the host(s) and other actors is appropriate.
There is a good relationship between the audio- and video tracks.
The programme is visually exciting.

The identification and the definition of educational objectives are the first, and probably the main tasks of the formative evaluator as objectives are the main yardsticks for measuring the educational merit of a programme (Mayer, 1980).

Mager (1975) and Scriven (1981) enrich the criteria for the evaluation of an educational product by suggesting an additional approach for selecting the criteria which Magar (1975) has termed 'reference criteria' and Scriven (1981) 'Goal-free evaluation'. "In this approach, the evaluator would make a deliberate attempt not to be co-opted by the rhetoric of programme goals" (Scriven, 1981:179). Accordingly, the evaluator uses the evaluation subject as a source of criteria, regardless of the stated goals, he identifies issues which exist in the educational programmes, and evaluates the importance of these effects for meeting educational needs.

Consequently it is the evaluator's role to rephrase the objectives as the programme progresses. The last stage of our model in which objectives can still be changed, is the stage labelled 'content'. If new objectives arise at the script stage, they should be examined but limited in number, or avoided to prevent a continuous development process.

Figure 6 is an example of how reference criteria are drawn from a script taken from episode 2 of the ETV-programme 'Time Management' (pictures 29-32 RTV1, SABC, 1984).

In this chapter, the bridged approach is used. Objectives from both sources are considered: objectives are stated clearly before the development of the programme, and are stated in the rationale of the programme. These objectives are inserted in the programme by its developers as 'reference criteria'.

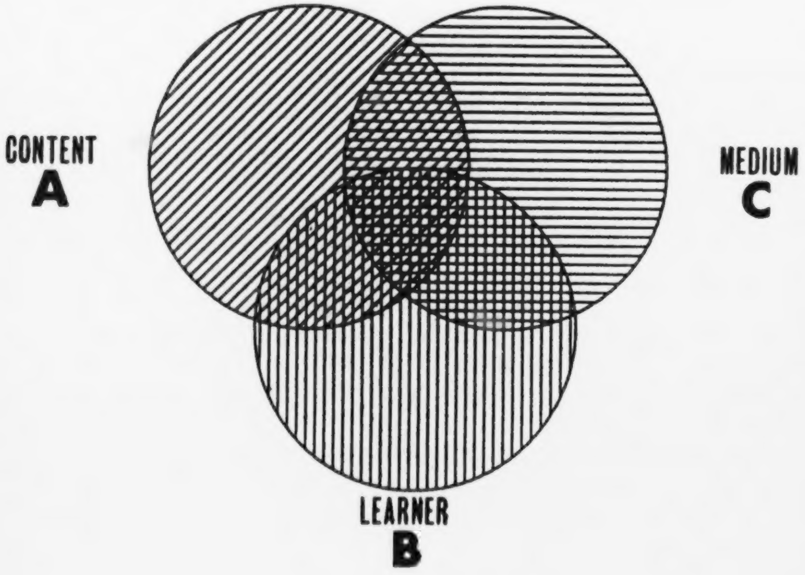
FIGURE 6: REFERENCE CRITERIA

PICTURE NUMBER	THE SCRIPT	REFERENCE CRITERIA
29	Some ways of overcoming procrastination will be shown: One way is "do it now" in order to overcome the habit of procrastinating.	At least two ways of showing how to overcome the procrastination habit.
30	Do it now - one has to ask oneself each 15 minutes ...	is it sufficient, too little or too much to ask oneself each 15 minutes?
31	The question: what is the best use of my time now?	Do pleasant words help to manage time?
32	One has to do now what has been postponed. One has to plan what to say before saying. One has to start telephone conversations with pleasant words.	(Preventing postponed action?)

Overview of Criteria for ETVP Evaluation

Three elements are joined to create an instructional TV situation: the content of the ETVP, TV features and learner-viewer characteristics. The three components are presented in Figure 7.

FIGURE 7: OVERLAP MODEL OF THREE COMPONENTS



Criteria for the analysis of ETVP will therefore stem from the three areas. The 'learner' component is further divided into cognitive and affective factors, and concern also the unique needs of particular learner groups.

The criteria for evaluating ETVP potential (the content) and viewer (the learner) abilities refer to two behavioural domains: the cognitive and the affective (Bloom, 1976; Guildford, 1959). Criteria for evaluating the medium stem from the same domains, as well as from the peculiar qualities of TV.

COGNITIVE CRITERIA

Cognitive criteria are clustered in two groups by Taba and Hills (1965): lower mental functions and higher mental functions. For major cases, this rough classification is adequate. In low mental activities, one can include knowledge and the recall of the information presented.

Bloom (1956) divides higher order mental functions into five categories according to the following hierarchy: comprehension of the message, the ability to translate it into other words or symbols and the ability to give examples to illustrate the message; application of the message to solve problems; analysis of the message in order to identify its components and the relationships between them; synthesis of the analysed components to create something new, based on these components; evaluation of the message according to standards.

Simultaneous Analysis of the ETVP and of the abilities of the target audience will provide the researcher with data concerning the suitability of the programme for the audience. Cognitive criteria from another point of view may be drawn from Piaget's theory (Piaget, 1952, 1957) on the development of intellect, which implies that the first thinking formats created in the human mind are elements of concrete thinking. One can refer to concrete thinking manipulation abilities in 2-5 year old children. The concrete thinking gradually develops into abstract thinking (which is the way adults think) via generalization of concrete elements in 6-9 year old children. The concrete-abstract elements in the programme should be analized and matched with the viewers' ability. This is why one can produce a TV programme aimed at pre-school children who might find a book too abstract to be a learning instrument (Mincowich, 1966).

AFFECTIVE CRITERIA

The power of TV as an instructional medium lies in its ability to change viewers' attitudes, thus criteria from that field should be employed in evaluation irrespective of whether the ETVP consists of such values, or whether they match the objectives of the instructor.

Bloom (1976) identified five categories in the affective domain according to the following hierarchy: receiving (attending and reacting); responding (active) valuing - commitment to certain values and sustained valuing - commitment to certain values and sustained interest; organising - looking for a conceptual explanation.

CRITERIA STEMMING FROM THE QUALITIES OF THE MEDIUM

Although, technically, TV production is a complicated process (Creaser, 1981; McNaught, 1979; Stephen, 1975) it is very easy for the viewer to operate and thus to explore as well. Almost every person has access to a TV-set, this makes almost everyone a potential member of an ETV target audience. As the audience is exposed to the TV for long periods (at a relatively low cost), TV provides the educationalist with an extra educational medium to be explored. (The viewer can be compelled to watch or can be a volunteer.)

Washee (1966) poses some critical questions on the effectiveness of TV programmes which can be used for the evaluation of programmes used in formal education, such as:

* will the ETVP be used for direct teaching or for enrichment only?

* what knowledge is needed of technical aspects of TV and TV equipment in order to be able to use the ETVP as an instructional tool?

* what kind of classroom climate will enhance the effectiveness of TV instructions?

In a new system, launched at the SABC, quality assessment, from the technical point of view, is the interest of a group of technical experts who evaluate the technical qualities of the picture and of the sound formatively (they test the pilot programme) and summatively (the final programme). This technical assessment provides important material for the evaluation team to work on, and saves the evaluator becoming involved in a type of assessment for which he is not qualified.

Mayer (1980) regards the viewers of a TV programme as a collection of individuals. In an experimental study, he found that 50 percent of a group of viewers did not understand a message transmitted to them in a TV advertisement and concluded that among viewers of the same programme there are a variety of levels of understanding. As the technical characteristics of a programme or its format influence the understanding of the individual viewer, these characteristics should be studied and evaluated (Baggaley, 1980b; Burke et al, 1981; Paluzzi, 1980). One should establish order in the process and group the criteria into a few categories. One way is to divide them into criteria which deal with technical aspects and are of interest mainly to the production team, and criteria which deal more with the content of the programme, and how it is shown in TV.

Goldevin (1976, 1980) sub-divided the second group. His three categories were:

- *Technical variables: i.e. camera angle, close-ups, setting, lighting effects, sound effects, selection of colours, etc. To these Burke et al (1981), added editing techniques such as cuts to linking shots, sequencing of shots, etc.*

- *Format: i.e. opening and closing, pacing (Is it in the opinion of the target audience, too fast, too slow, or correct Baggaley - 1980a)?, built-in repetitions (Rydin, 1973), relationships between verbal and visual density of information, Burke et al, (1981) and Salomon and Snow (1968) (Is the educational message salient)?*

- *Performing variations: Performer's age, gender, eye contact level, etc.*

Only a few researchers have studied the influence of the qualities of the medium on the viewer's cognitive skills. Salomon's study is a leader in the field.

Salomon (1976:25) found "at least three factors which could make a difference in learning: the technology involved, the content transmitted, and the symbolic codes into which the message is coded": he conceptualised TV as having its own symbolic codes. The learner must familiarize himself with these TV codes in order to 'translate' the message transmitted into a meaningful cognitive message because the knowledge acquired by each use of a different code is different in amount and content. He found, for example, "that children viewing the version which emphasized close-ups acquired more knowledge pertaining to the relations between parts of wholes" (p 26).

The close-up technique, which is a unique technique used often on the 'small screen' can develop the cognitive skills of analysis and differentiation in the viewer if planned, or it can be used in other ways to emphasize things. Therefore, one should test an educational programme on issues such as: does the close-up emphasize an important element (from the educational point of view) or distract the viewer's attention to a side issue? If the differentiation between two elements is the educational objective of a programme - is the close-up in the programme used to help the viewer to distinguish between the two?

Salomon found that "children viewing the version which entailed logical gaps, excelled in their comprehension of the plot's logical structure and continuity" (1976:26). Such planned and unplanned gaps can serve as educational criteria as well. Other criteria of the same nature are the use of humour in the programme, dramatic effects, use of colours (Lee and Schromm, 1967; Murray, 1980; Clements, 1984) the rhythm of the programme, etc.

Lesser (1977) and Cook and Curtin (1985) pointed out that the 'Sesame Street' programmes which imitated the pace of American TV advertisements probably caused hyperactivity, short attention spans and lack of interest among young viewers in later slower-pace class activities. This crucial finding was heavily criticised (Anderson, 1985), yet it can suggest a very important criterion concerning the pace of a programme for a specific target audience, as several studies "have results suggestive that specific subgroups of children may be influenced more than others" (Anderson, 1985).

But since "there is relatively little research on the effects of television viewing on children's attentional abilities" (Anderson, 1985) and as "where effects and associations with attentional skills have been found, none of the studies have identified what aspects of television may be responsible. Despite hypotheses concerning the negative influence of rapid pacing, for example, only one study directly examined pacing. This study found no effect". (Anderson, 1985).

It can be concluded that the pace of a programme can be tested with a target audience in order to establish acceptability - but no criterion has been established for the optimal pacing of TV programmes.

Anderson (1985) also discusses violence in children's programmes and concludes that "there is not enough research yet to conclude that violent action programming actually causes negative changes in attentional behaviours". The conclusions are that "television may indeed influence children's attentional skills, but if so, the influence appears to be positive as well as negative".

What to evaluate, when and by whom? One can phrase the above criteria in question format which is a convenient way of presenting the problem to an appropriate expert (Gall, 1981) e.g. Do the children like the colour setting of the scene? (pictorial). Can you identify a possible negative influence of an aggresive behaviour in the programme? (for an expert in child psychology). Does the music on the sound track complement the message, suit it or interfere with it? (sound). Is the plot funny in the opinion of the target audience? Is humour used in the programmes? Is the camera angle used for the presenter a favourable angle for him?

"One of TV's major advantages is its capacity to overtly and explicitly show a process, a transformation ... TV alone can visually show a transformation which is analogous, or even similar to what ought to take place in our minds" (Salomon, 1976). The evaluator can therefore examine, in the rationale stage, whether the programme is about a process or a static situation and can decide accordingly about what to show in the TV programme and what to leave for books or slides.

(Criteria related to the medium domain, are mentioned in the early stages of the ETVP development, and thus can also be analysed in the early non-experimental stages.

The main role of the evaluator is to identify such issues, to analyse them and their effects, and draw the manager's attention to the conclusions; and to follow up the interpretation of the recommendations by the production team and see that they are correctly implemented. As it is difficult to predict how the sound effects, camera angles, etc. will effect the viewers before they can actually see it on screen, the evaluation of these should occur only at the 'last moment' of formative evaluation, namely with the pilot programme or pilot segment. Nevertheless, criteria related to the medium domain are mentioned in the early stages of ETV development and thus can be analysed in the early non-experimental stages as well.

Time and money are pressuring factors and the production team is likely to be very unco-operative at this stage. It is the task of the evaluator to extract as many of the criteria for examination as possible by testing them beforehand, and by deferring to later stages the evaluation of production issues and other issues that can not be examined in early developmental stages.

The Role of the Formative Evaluator of ETV

The criteria have become subjective and flexible and measurements are plainly dependent on the technical ability of the producer. The evaluator's technical understanding has progressed to such an extent that he is able to relate an educational outcome to a technical quality.

In order to accommodate the educationalist and the TV producer in the self-formative evaluation situation, they have to be encouraged to learn each other's fields. The educationalist must learn production techniques and the producer must study educational concepts.

By shifting the main emphasis of the formative evaluation from the testing of the actual programme with the target audience to the evaluation of the content of the programme and its values, and by increasing the importance given to educational moral and scientific criteria, one can hope that the formative evaluation, in the early stages, will not only increase the effectiveness of the evaluation techniques, but will also help increase the educational scope of our TV programmes.

A model which presents the role of the formative evaluator and the criteria to be considered at each stage is described in the following figure. Activities which have a star (*) can be self-evaluation activities.

FIGURE 8: FORMATIVE EVALUATION OF ETV DURING THE DEVELOPMENT STAGES

STAGE	FORMATIVE EVALUATION ACTIVITIES	CRITERIA	PERSONS INVOLVED
Idea	Collecting available information Needs Assessment Selecting goals* Identifying and defining audiences* Compiling internal reports	Needs Goals: Educational, Natio-Scientific etc.	Researchers (external and Internal) Experts
Rationale	Target Audience Analysis* Medium Analysis* Collecting and Analysing Exp. Judgement Compiling Internal Reports	Goals Selected Needs Objectives Target Audience Features: Cognitive, Affective and Psychomotoric	Manager, Writer, Internal Evaluator, Experts
Content	Collecting and Analysing Experts' Judgement and Opinions* Matching with Rationale* Leading Feedback Sessions* Compiling Tentative Reports	Objectives Reference Criteria Scientific and Educational Issues Socio-psychological Variables	Internal Evaluator, Writer, Experts
Script	Final Phrasing of the objectives* Matching with content* Collecting Information from Potential Users Leading Feedback Sessions*: Interpreting Recommendations Compiling Reports	Objectives Reference Criteria Predicted Production Variables: Codes, Repetitions Links, Sequences Educational Input Entertainment Issues	Internal Evaluator Content Writer Script Writer (Producer) Experts in the Subject Matter, Behavioural Sciences and Production
Pilot-Programme	Collecting Information from Viewers Analysing the pilot-programme*: viewers, medium and content* Compiling and submitting a written report	Objectives: Acceptance by Target Audience Production Qualities: Technical and Organizational Performance Editing	Internal and/or External Evaluator Experts Producer Manager (Script Writer)
Design of the Final Programme	Interpreting recommendations*	Evaluation Recommendations	Evaluator Leader* Manager

REFERENCES

Abedor, A J 1972. Development and field of a model of formative evaluation. Viewpoint, 48:9-43.

Anderson, D R 1985. The Influence of Television on children's attentional abilities. Amherdt, University of Massachusetts, MA, December, 1985

Alkin, M C 1974. Evaluation and Decision-Making: The Title VII Experience. Los Angeles, University of California, Center for Study and Evaluation.

Anastasi, A 1976. Psychological Testing. New York, MacMillan, (4th Ed.).

Aversa, F M and Forman, D C 1978. Issues in evaluation of educational television programmes. NSPI Journal, March, p 16-18.

Baggaley, J P 1980a. Psychology of the TV Image. Westmead, Gower.

Baggaley, J P (ed.) 1980b. Proceeding of the Third International Conference on Experimental Research in Televised Instruction.

Baggaley, J P and Duck, S W 1975. Communication effectiveness in the educational media: three experiments. J Baggaley, G H Jourisson and H Marchant (eds.), Communication and Learning, Aspects of Educational Technology, Vol. 8, London, Pitman.

Baggaley, J P and Duck, S W 1976. Dynamics of Television. Hampshire, Gower.

Ball, S and Bogatz, G A 1970. The First Year of Sesame Street: An Evaluation. Princeton, Educational Testing Service.

Ball, S and Bogatz, G S 1971. The Second Year of Sesame Street: A Continuing Evaluation. Princeton, Educational Testing Service.

Bloom, B S 1956. Taxomony of Educational Objectives: The Classification of Educational Goals. (Handbook 1: Cognitive Domain). New York, Longman, Green and Co.

Bloom, B S 1976. Human Characteristics and School Learning. New York, McGraw-Hill.

Burke, N C, Dowaliby, F J and McKee, B G 1981. A Film Evaluation Checklist to Describe Instructionally Relevant Film Characteristics: a Rater Reliability Study. Department of Educational Research and Development, National Technical Institute for the Deaf, Paper Series No. 42.

Chu, G C and Schramm, W 1967. Learning from Television: What the Research Says. Washington D.C.. National Association of Educational Broadcasters.

Clements, D H 1984. Implication of media research for the instructional application of computers with young children. Educational Technology, Vol. 24, 7-16.

Coldevin, G O 1976. Comparitive effectiveness of TV production variables. Journal of Educational Television and Other Media, 2(3), 87-92.

Cook, T D, Appleton, H, Conner, R F, Shaffer, A, Tamkin, G and Weber, S J 1975. "Sesame Street" Revisited. New York, Russel Sage.

Cook, T D and Curtin, T R 1985. ERIC/ECTJ annual review paper, evaluating the CTW model for producing educational television, ECTJ, 33(2), 91-112 Washington D.C.

Creaser, H 1981. The impact of media on learning. Journal of Educational Television and Other Media, 17(3), 99-101.

Cronbach, L J 1963. Course improvement through evaluation. Teachers' College Record, 64(5), 672-683.

Domingo, A 1979. Classifying the Concept of Educational Objectives. Lisbon, Groupo dé travolko Posto o Ensi no da Biologica (mimegraph).

Duby, A 1985a. Formative evaluation of educational television. P J Van Zyl, L L Breytenbach, A Duby and J Grove (eds.), Educational Programmes RTV1. Johannesburg, SABC, 25-34.

Duby, A 1985b. Self-formative Evaluation of Instructional Materials with Reference to Games and Television. Doctoral Thesis, University of South Africa, Pretoria.

Duby, A 1985c. Executive report: Exploratory Study for the Formative Evaluation of the ETV Programme "Ziki Zikombot". Johannesburg, ERTV1, SABC (Unpublished paper).

Duby, A 1986. Self-Formative Evaluation of Instructional Materials. Paper presented at the ICA International Communication Association, Chicago, 22-26 May.

Duby, A 1987. Formative Evaluation of ETV Programmes in Early Stages of their Development, Paper submitted to the AERA annual conference, Washington D.C. April 1987.

Eisner, E U 1979. The Educational Imagination: The Design and Evaluation of Social Programmes. New York, MacMillan.

Fox, S 1971. A Practical Image of the Practical. Paper delivered at the annual meeting of the American Educational Research Association Convention, New York.

Fox, S 1985. The vitality of theory in Schwab's conception of the practical. Curriculum Inquiry, 15(1), 63-90.

Gall, M D 1981. Handbook for Evaluating and Selecting Curriculum Materials. Newton, Allyn and Bacon Inc.

Guilford, J P 1959. Three faces of the intellect. The American Psychologist, 17, 469-479.

"Harry's House" 1985. SABC, Educational TV programme, episode 1.

Land, H W 1972. The Children's Television Workshop: How and Why it works. Jericho, New York. Nassau Board of Co-operative Educational Services.

Leide, G 1977. Expert judgement as evaluation data. A Lewy (ed). Handbook of Curriculum Evaluation. New York, Longman.

Lesser, G S 1974. Children and Television: Lessons from Sesame Street. New York, Vintage Books.

Lewy, A 1971. The Practice of Curriculum Evaluation. Jerusalem, Ministry of Education, Curriculum Centre.

Lewy, A 1973. Utilising experts' judgements in the process of curriculum evaluation. California, Centre for the Study of Evaluation UCLA Graduate School of Education, Report 87(7).

Lewy, A 1977. Handbook of Curriculum Evaluation. Paris, New York, Unesco.

Lewy, A 1985a. Formative and summative. T Husen and T N Postlethwaite (eds.), International Encyclopeadia of Education. Oxford, Pergamon Press.

Lewy, A 1985b. The autonomous unit of evaluation: Combining the strengths of in-house and external evaluations. P Tamir (ed.). The Role of Evaluators in Curriculum Development. Beckenham, Croom Helm.

Light, J A and Reynolds, L J 1972. Procedures for formative evaluation. Viewpoint, 48:45-47.

Mager, R F 1975. Preparing Instructional objectives. Second edition USA. F Caron Publishers Inc.

Mann, J S 1969. Curriculum criticism. Teachers College Record, 71, 27-70.

Markle, S M 1967. Empiricial testing of programs. P C Lange (ed.). Programmed Instructions, The Sixty Sixth yearbook of National Society for the Study of Education Part II. Chicago, NNSE.

Mates, B F 1980. Current emphases and issues in plannes programming for children. E L Palmer and A Dorr (eds.). Children and the Faces of Television: Teaching, Violence, Selling. New York, Academic Press, 19-32.

Mayer, M 1980. The comprehension problem. American Film, 9:8-9.

Mielke, K W 1968. Questionning the ETV questions. Educational Broadcasting Preview, (2)6.

Mielke, K W and Chen, M 1981. Children television, and Science: an overview of the Formative Research for 3-2-1 contact. New York, Children's Television Workshop.

McNaught, P C 1979. Educational television, 1969-1979. Journal of Educational Television and Other Media, 5(1), 4-5.

Mielke, K W and Chen, M 1983. Formative research for 3-2-1 contact: methods and insights. J A Howe (ed.). Learning from Television Psychological and Educational Research. London, 31-55.

Minkowich, A 1966. Experience and abstract thinking, Megamot, 14(4), 172-199.

Nevo, D 1983. The conceptualisation of educational evaluation. An analytical review of the literature. Review of Educational Research, 53(1), 117-128.

Nevo, D 1985. Experts' Opinion: A Powerful Evaluation tool. Paper presented at the annual meeting of the American Educational Research Association, Chicago, March 31 to April 4, 1985.

Nickerson, R B 1979. Programme Evaluation using the Programme Education Analysis Computer (PEAC). Paper presented at the 1979 convention of the Association for Educational Communications and Technology, New Orleans, March 1979.

Palmer, E L 1974. Formative research in the production of television for children. D Olsen (ed.). Media and Symbols: The Form of Expression, Communication and Education. The Seventy-third yearbook of the National Society for the Study of Education. Chicago, University of Chicago Press.

Palmer, E L 1978. A Pedagogical Analysis of Recurrent Formats on Sesame Street and Electric Company. Paper presented at the International Conference on Children's Television, Amsterdam.

Paluzzi, J V 1980. Interest and presentational formats. J Baggaley (ed). Proceedings of the Third International Conference on Experimental Research in Television Instruction. Memorial University of Newfoundland, p 22-35.

Piaget, J 1952. The Origins of Intelligence in Children. New York, University Press.

Piaget, J 1957. Logic and Psychology. New York, Basic Books.

Provus, M 1977. Discrepancy Evaluation. Barkely, McCuthan.

Rydin, I 1973. Information Processes in Pre-school Children. Stockholm Swedish Broadcasting Corporation.

Salomon, G 1974. Interaction of Media, Cognition and Learning. California, Gossey-Bass Publishers.

Salomon, G 1976. A cognitive approach to media. Educational Technology, (5), 25-27.

Scriven, M 1967. The methodology of evaluation. R W Tyler, R M Gagne and M Scriven (eds.). Perspectives of Curriculum Evaluation. Chicago, Rand McNally. AERA monograph series on curriculum evaluation, 1, 39-83.

Scriven, M 1974. Standards for the evaluation of educational programmes and products. G D Borich (ed). Evaluating Educational Programmes and Products. Englewood Cliffs, Educational Technology Publications, 5-26.

Scriven, M 1981. Evaluation Third Edition Thesaurus. California, Edgepress.

Stake, R E 1967. The countenance of educational evaluation. Teachers College Record, 68, 523-540.

Stephen, K B 1975. A matrix for quality, Journal of Educational TV and Other Media, 11, 2-6.

"Study Methods" 1985 SABC. Educational programmes RTV1, Episode 1. (ETV Programme).

Suarez, T M 1981. Needs assessment. A Lewy and D Nevo (eds.). Evaluation Roles. Gordon and Breach, 487-505.

Taba, H and Hills, J L 1965. Teacher Handbook for Contracosts Social Guides, Grades 1-6. Hyward, CA., Rapid Printers.

Tidhar, C 1979. Thirteen Years of Evaluation in Instructional Television Centre in Israel. Tel-Aviv, ITV.

Tyler, R W 1950. Basic Principles of Curriculum and Instruction. Chicago, Illinois, University of Chicago Press.

Van Vuuren, D P 1985. A formative evaluation of an educational programme: "The Edge". P J van Zyl, L L Breytenbach, A Duby and J Grove. Educational Programmes RTV1, Johannesburg, SABC, (1)1, 64-74.

Wasche, G A 1966. Utilising Television in the Classroom. Menlo Park, C.A., Pacific Coast Publishers.

Werner, W 1980. Editorial criticism in curriculum analysis. Curriculum Inquiry, 10, 143-154.

Ziki-Zikombot 1985. SABC Educational TV Programme.

GUIDELINES FOR EVALUATING ETV-PROGRAMMES FOR TEACHER TRAINING

N J de Jager and R H Young

General Introduction

"Command of knowledge is the ability to use it to make decisions, to solve problems, to answer questions or to nourish the process of reflective thought" (Ebel, 1974:127). Command of knowledge is therefore the key to true self-realisation. The most important task of the teacher is to increase the child's knowledge to such an extent that he is able to play a satisfying role in life. The explosion of knowledge and the resultant heightened aspirations and expectations of our time, have increased, to a great extent, the task of the teacher. Cropley (1977:12) for this reason emphasizes "... the need for a new educational goal ... education for a changing world".

Teacher-training has to keep pace with the increasing demands and responsibilities of our time. Attention must be given to revising the curriculum of trainee teachers, and to upgrading in-service teachers.

Modern media, especially educational television, offer exceptional possibilities for training. "We are in the 'century of the image'" (Fauquet, 1971:16). McDaniel (1982:16) was thinking along these lines when he emphasised the importance of television in education. In addition, Reed (1976:117) has the following comment to make on teacher training: "The best teaching should be done in the teaching of teachers. Television is the brightest new star in the teaching sky - it must not be ignored".

Educational television is one of the most important teaching aids in teacher training, especially for distance education. The role of educational television becomes meaningful when measured against the background of its objectives.

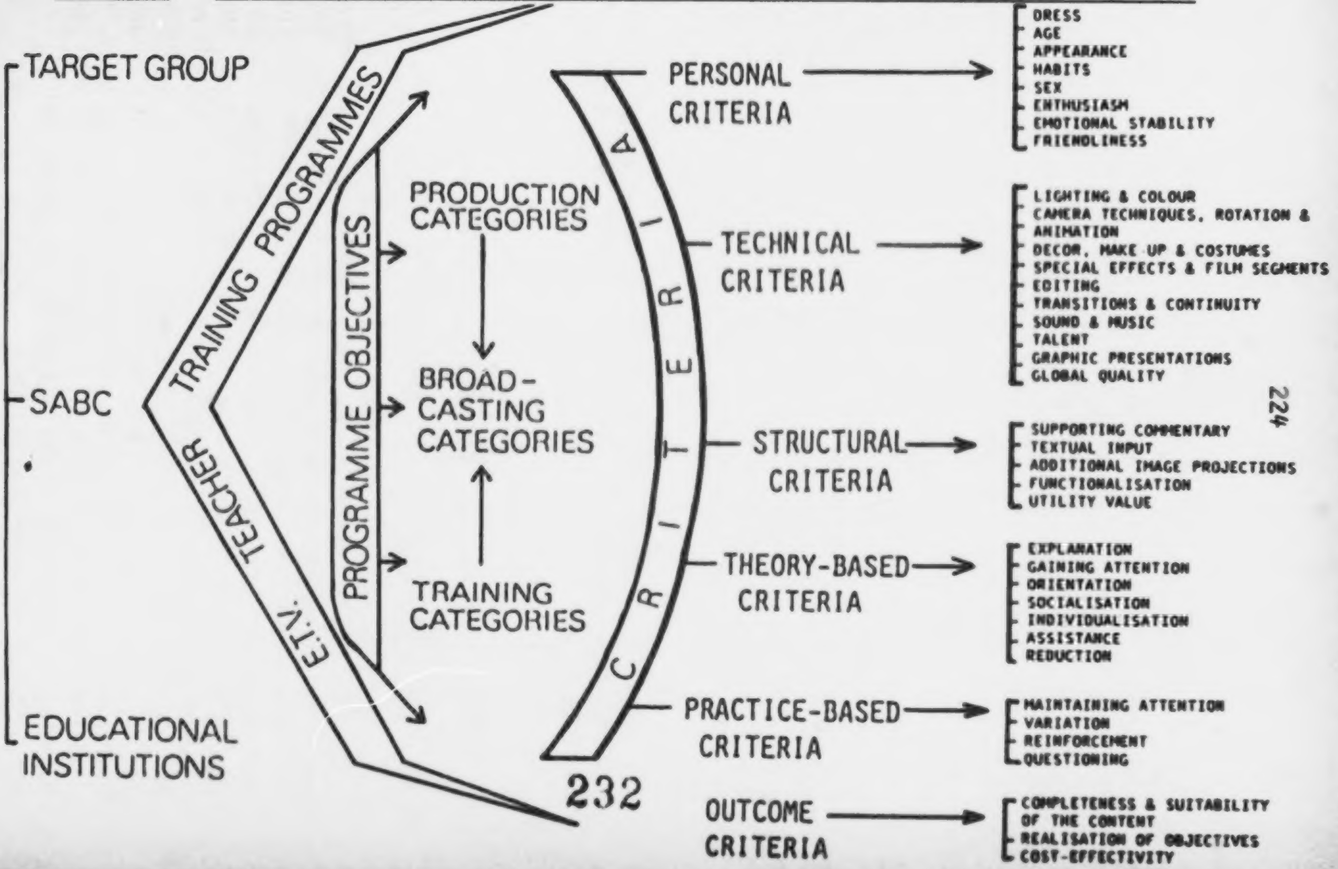
Bates (1984:18-32) classifies the objectives of educational television under three headings: expansion of knowledge, enrichment, and substitution. Educational television is a primary source of information, but it is also used to enrich existing knowledge. Enrichment entails the implementation of television to expand upon, and to reinforce certain facts and skills being conveyed to learners. It can make a teacher training programme more relevant, interesting and illustrative. Educational television can be used to introduce the latest research on teaching to a target group. Thirdly, educational television can be used as a teacher substitute, and for self-teaching. It can be used to substitute for teachers who are not available because of illness, resignation or retirement. In such cases, educational television programmes can be used as a supplementary aid or they can substitute for the teacher directly. The basic content and the style or method of teaching are embedded in the direct-teaching programme.

A criteria model (Figure 1) for evaluating educational television programmes for establishing teacher training needs is discussed below.

A Criteria Model for Evaluating Educational Television for Teacher Training

The model includes all the aspects of programme evaluation from production to application to teacher training. The criteria for evaluating the production aspects of the programme are applied before broadcasting, whereas criteria for evaluating the broadcast are applied at the time of the broadcast. The criteria for evaluating the training will only be applied after the broadcast. The aim will be to evaluate whether the content and skill objectives of the educational programmes have been met. Certain objectives may be emphasized more than others, and certain criteria may be ignored; the following model (Figure 1) identifies, inter alia, those criteria relevant to evaluation.

FIGURE 1: A CRITERIA MODEL FOR EVALUATING ETV-PROGRAMMES FOR TEACHER TRAINING


TARGET GROUP
SABC
EDUCATIONAL INSTITUTIONS
E.T.V. TEACHER TRAINING PROGRAMMES
PROGRAMME OBJECTIVES
PRODUCTION CATEGORIES
BROAD-CASTING CATEGORIES
TRAINING CATEGORIES
CRITERIA
PERSONAL CRITERIA
DRESS
AGE
APPEARANCE
HABITS
SEX
ENTHUSIASM
EMOTIONAL STABILITY
FRIENDLINESS
TECHNICAL CRITERIA
LIGHTING & COLOUR
CAMERA TECHNIQUES, ROTATION & ANIMATION
DECOR, MAKE UP & COSTUMES
SPECIAL EFFECTS & FILM SEGMENTS
EDITING
TRANSITIONS & CONTINUITY
SOUND & MUSIC
TALENT
GRAPHIC PRESENTATIONS
GLOBAL QUALITY
STRUCTURAL CRITERIA
SUPPORTING COMMENTARY
TEXTUAL INPUT
ADDITIONAL IMAGE PROJECTIONS
FUNCTIONALISATION
UTILITY VALUE
THEORY-BASED CRITERIA
EXPLANATION
GAINING ATTENTION
ORIENTATION
SOCIALISATION
INDIVIDUALISATION
ASSISTANCE
REDUCTION
PRACTICE-BASED CRITERIA
MAINTAINING ATTENTION
VARIATION
REINFORCEMENT
QUESTIONING
OUTCOME CRITERIA
COMPLETENESS & SUITABILITY OF THE CONTENT
REALISATION OF OBJECTIVES
COST-EFFECTIVITY

THE EVALUATORS

The Target Group

Although the criteria are aimed mainly at evaluating in-service programmes for serving teachers, closed-circuit television can also be used for the initial training of teachers. The target group for the evaluation of the teacher training programmes will thus be in-service teachers as well as student teachers.

South Africa's teachers operate in first- and third-world situations, consequently their needs are likely to differ, and their interpretations of the evaluation criteria may also differ. Personal criteria used in evaluation are open to different interpretations and emphases if teachers' philosophies of life differ.

The Broadcasting Corporation

The broadcasting corporation is responsible for the production and transmission of programmes. Process evaluation will be undertaken during production and transmission to obviate problems before the programme is finally transmitted. It is expected that the evaluation will be undertaken by the programme director and his staff, for whom technical quality will be of special importance, and by the scriptwriter and committee of subject specialists, for whom the realisation of objectives will be the highest priority. The broadcasting corporation will, however, also be interested in product evaluation because the achievement of objectives will depend largely on the length of the broadcast, the time and day of broadcast, and the like.

Educational Institutions

Professional and formal training institutions tend to respond to educational television programmes for teacher training from a professional and educational perspective. Evaluators from this field are likely to require product evaluations as they are interested mainly in outcomes. Criteria for evaluating learning outcomes are thus important.

Yet the success of the programme depends on a variety of factors. The programme may have failed because of errors in the production stage or because it was transmitted at the wrong time. The final evaluation will thus have to encompass all aspects of the programme (see Figure 1). Evaluation by education specialists is of the utmost importance to the producers as well as the transmitters of programmes.

Objectives and the Components of an Educational Programme

The objectives will determine the nature of the educational programme. One programme may concentrate on subject content while another may focus on subject methodology or verbal interaction in the classroom. For each aspect different criteria of the evaluation model will be emphasized. Certain criteria may not be relevant to certain programmes.

Educational programmes for teacher training have two major components; production and training (see Figure 2). Broadcasting authorities need to be aware of these two components when planning teacher training programmes.

FIGURE 2

PRODUCTION → *TEACHER TRAINING PROGRAMME* ← *TRAINING*

The production component is concerned with technical and technological aspects of television, and in particular with those factors that are necessary for compiling an educational television programme. The training component is oriented towards education and training, and as such towards factors that may affect the academic and professional aspects of the teacher or teacher trainee.

A programme that is pedagogically unsound may fail even though the technical and technological aspects function smoothly.

Criteria for the Evaluation of the Production Component of an Educational Programme

The production component of educational television comprises three facets: presentation, structure, and technical approach. Although there are different evaluation criteria for each of these production variables, there are also criteria common of them all. Despite the differences, the variables cannot be evaluated separately. The whole picture must be evaluated. Breytenbach (SABC)[1] refers to this total picture as the 'picture on the screen'.

A brief discussion of personal, technical and structural criteria as factors in production for the evaluation of the 'picture on the screen' follows.

PERSONAL CRITERIA

The presenter plays an important role in making an educational television programme effective. Schramm (1972:16), referring to his previous experience in compiling programmes for educational television, claims that programmes that are very soundly planned and structured can be totally unsuccessful if the wrong presenters are chosen.

How can the presenter of teacher training programmes achieve the desired effect? It is extremely difficult to draw up a set of specific criteria for evaluating a person's conduct and behaviour. Grové (SABC)[2] claims the presenter's behaviour and appearance should encourage the teacher or trainee to identify with him.

---

1. Personal interview with Mr Breytenbach, Researcher, ERTV1, SABC July 1985.

2. Personal interview with Mr Grové, Manager, ERTV1, SABC July 1985.

If the presenter is required to play the part of a teacher, his dress, appearance and even demeanour should not detract from the status of a teacher. "He is viewed first and foremost as a stimulative object which must arouse in the student's positive emotional responses toward learning" (Miron and Segal, 1978:28).

TECHNICAL CRITERIA

For a television-conscious community, high technical quality of the programme is important, especially if the programme is used for training. "Those who are spoiled by having access to first-class shows in their leisure time will frown on ETV-programmes that are considerably under the technical standard of such shows" (Schramm, 1977:7). Teacher training programmes are often used to demonstrate teaching strategies and subject methodology. In such cases the programmes are likely to be less effective if the technical quality is poor.

Shively (1975:51) views the technical quality of a television programme as a matter of professional judgement, based on broadcasting standards and technical expertise. To evaluate educational television programmes adequately, both production specialists and educationalists with practical experience and technical expertise should be consulted. Educationists, will be required to evaluate teacher training programmes, but technical proficiency can also contribute to making programmes acceptable educationally.

The following criteria can be used for evaluating the technical quality of television programmes. Different emphases will be placed on programmes for teacher training. The differences in emphases are discussed later.

- Colour - *This criterion comprises factors such as brightness, intensity, balance and the use of colour in the programme.*

- Lighting - *The amount of light and its direction, quality, balance and specialised use.*

- Cameras - *The composition of the shots, the movement of the cameras, the sequence of the shots, the correct use of close-ups, slow motion and zoom.*

- Rotation - *Making objects rotate on the screen in order to create a three-dimensional image.*

- Animation - *Drawing and movement as components of animation.*

- Sets - *The functional use of, and styles and types of, sets.*

- Make-up and costumes - *Their appropriateness and style.*

- Special effects - *Special effects such as inserts can enrich a programme if they are used with discretion. Special effects can be classified as optical, chemical, mechanical or electronic.*

- Film inserts - *The way in which film is used as an insert in a programme or as part of the background.*

- Editing - *This criterion deals with the quality and style of editing.*

- Transitions - *Changing from one camera shot to another requires cutting, fading and dissolving.*

- Continuity - *How the flow of the programme maintains interest.*

- Sound - *Includes the placing and use of microphones, balance, the level of foreign sounds, the mixing of sound and synchronization.*

* Music - *The selection, theme, arrangement and appropriateness of music.*

* Talent - *The central focus of a television programme is often the actors and presenters. This criterion refers to the quality of the moves, delivery and voice of the central and supporting actors/presenters.*

* Graphics - *The quality of maps, graphs, charts and drawings used in the programme.*

* Overall quality - *This criterion allows for an overall general judgement of the technical quality of a programme.*

(Shively - 1975:49-50)

STRUCTURAL CRITERIA

Besides production and training criteria for television, there are additional criteria for evaluating the actual broadcasts of the programmes. The following are some of these criteria which deal basically with the structuring of a programme to obtain a specific effect with the 'picture on the screen'.

* Supporting commentary - *Aside from the ordinary oral interaction, it is sometimes necessary to use a supporting commentary to focus attention on important aspects of the programme. Points of great didactic value can be emphasized in this way.*

* Textual input - *Where supporting commentary disturbs the flow of a programme, explanatory textual input can be provided on the screen. Such projections have to be used judiciously as they may disrupt viewing.*

* Additional image projections - *Additional images can be projected onto the screen from time to time to highlight the theme for the target audience, or to provide additional information. Historical events can be brought to life using such projections without interrupting the flow of the programme. Care must be taken not to disturb the flow of the programme or lower its quality. The so-called 'talking head' can sometimes be effective but over-use of this devise can make the programme boring.*

* Background music - *Music, if it is used effectively, increases the impact of a programme. In training programmes of a professional nature, care must be taken with the choice of the background music.*

* Function - *The target audience can be involved actively in the training programme by getting them to research some matter related to the theme or by answering of questions. A supplementary guide containing questions, assignments or tests can be used to link the training programmes to courses at teaching centres or holiday schools. Assignments must be functional and relevant to involve the teachers.*

* Utility value - *Teacher training programmes may also be watched by other members of the viewing public. These programmes should appeal to teachers as well as casual viewers.*

Criteria for Evaluating the Training Component of Educational Programmes

There are diverse views on the nature and content of teacher training programmes and on the principles of good teaching. Ornstein (1976:202) says in this regard: "Most educational researchers admit that 'they do not know how to define, prepare for, or measure teacher competence'".

It is difficult to establish effective criteria for evaluating teacher training. Evaluation criteria fall into two categories. The first set of criteria can be used to determine the extent to which the training programme complies with established didactic theory at the tertiary level. The second set of criteria is more practically orientated and places the focus on the behaviour of the teacher in the classroom.

THEORETICALLY BASED CRITERIA FOR EVALUATION

Teacher training programmes should conform to established didactic theories; consequently theoretically based criteria should also be used to evaluate training programmes. Some of the didactic categories of van der Stoep and Louw (1976:40-49) have been used to identify relevant evaluation criteria. The following criteria have been derived from didactic categories (Mobbs, 1984:58):

Didactic Categories	*Evaluation Criteria*
Interpretation of reality	*Explanation*
Learning	*Holding attention*
Orientation	*Orientation*
Socialization	*Socialization*
Individualization	*Individualization*
Support	*Assistance*
Reduction	*Reduction*

Although the above criteria are not comprehensive, they can be used to determine whether training programmes comply with didactic requirements.

Explanation - Teaching is the process in which unknown or unclear material is presented to a learner in such a way that it becomes accessible to him. ETV has the potential for self-teaching, and could be used to demonstrate certain materials, skills, insights, relations and principles of teacher training.

The following are some of the special techniques for explaining (Wessels, 1982:85):

- *Time-lapse photography. Slow movements such as the germination of a seed or the opening of a flower are speeded up to portray the whole process in a very short space of time. This particular aspect of television can be of great value in teaching subject disciplines, e.g. the teaching of biology.*

- *The slowing down of movement. The path of a fast-moving object can be slowed down for training sports instructors so the pole-vaulter's movements can be studied as he moves over the cross-bar or a tennis player's service can be analysed because it is possible to control the motion captured by the camera.*

- *Micro-photography. Objects not visible to the human eye can be enlarged to such an extent that they are seen. In this way the biology teacher can look at micro-organisms and explain cell division.*

- *The zoom technique. Objects can be viewed first in the context of a larger whole and then brought nearer for a study of the details. A person can be shown on the screen as part of a crowd and can then be zoomed nearer for close observation. Training programmes with an historical theme can make good use of this technique.*

Holding attention - A television programme for teacher training should promote learning. Learning is facilitated when attention is held. The attention of the target group should be focused constantly on the material in question. This criterion will be discussed in more detail under the heading 'Practice-based criteria'.

Orientation - Orientation as a didactic criterion relates to the fact that while he is learning, the learner determines his own position in relation to the material being studied so as to understand reality better (Mobbs, 1984:63). Using orientation as his measure, the evaluator can determine the extent to which the training programme has provided the learners with experiences of historical events or distant places. The past can be reconstructed, and events in other places and at other times can be actualised.

Socialization - An effective training situation is, in essence, a situation in which the teacher and the learners meet and interact with one another. Using socialization as criterion, the evaluator can determine the extent to which the training programme has established interrelationships and promoted co-operation inter se. Programmes based on small group discussions and team-teaching to promote group relationships can be used.

Individualization - People's individuality cannot be denied, nor should it be overlooked. In education we expect learners with different abilities to be treated differently by the compiler of ETV-programmes. Differentiation can be achieved in ETV training programmes with self-learning.

Assistance - Assistance or support is a basic characteristic of all learning and training situations. All training programmes should therefore be measured against this criterion. At the planning stage, the programme compilers should determine, by formulating their objectives, the degree to which the target group should receive assistance with certain skills or strategies. Concentration on the visual possibilities of training programmes sometimes results in the neglect of follow-up assistance, e.g. pamphlets, directives, etc.

Reduction - Because of time and costs, it is important that unnecessary and irrelevant details be left out of television programmes. Only the essential elements necessary to reconstitute reality should be used and emphasized to such an extent that a better understanding becomes possible.

PRACTICE-BASED CRITERIA

Criteria that are relevant to promoting practice, focus on the proficiency of the teacher. The assumption is that the behaviour and conduct of a teacher in a classroom situation can make the teaching more effective. Brusling (1974:16) claims "No-one can entertain a doubt that there is a mastery in teaching as in every other art. Nor is it less obvious that within reasonable limits this skill and this mastery may themselves be made the subject of instruction and be communicated to others". This statement nullifies the widely held belief that good teachers are born to teach.

Structuring of training programmes becomes thus a more meaningful activity. Brown (1975:5) expresses himself quite unambiguously: "Teaching also has its repertoire of skills. If this cannot be identified and taught then many teacher-educators are guilty of a huge confidence trick".

Not all teacher training television programmes focus necessarily on teaching skills nor do all skills figure equally prominently in all teaching situations: yet the following can be suggested as criteria for evaluating the effectiveness of teacher behaviour:

- *Gaining attention* - *Learning will not be effected without gaining the learner's attention. It is, therefore, the task of every teacher at the start of a lesson unit to act in such a way that he gains the full attention of the learners. This is also true for training programmes. The action in the programme presentation must stimulate the curiosity of the learners so that they will want to know more about what is to follow. The context of the target group must, however, be born in mind. Activities which may fascinate primary school teachers will not necessarily gain the attention of high school teachers.*

- *Variation - The learner's attention should be gained, but if learning is to take place, sustained attention is equally important, and variation in the elements making up the situation is therefore extremely important throughout the presentation. The following can serve as sub-criteria for variation:*

    - Use of the voice - This refers to variation in tempo, tone and volume.

    - Facial expression - It is important that eye contact be maintained and that facial expressions be appropriate so that they support the verbal interaction.

    - Body movement - Non-verbal interaction should give support to verbal communication throughout. Hand and other movements must be appropriate, as exaggerated movement can be distracting.

    - Method of presentation - The methods of presentation should be constantly varied: there should even be variation in respect of the type of media used.

- *Reinforcement - Reinforcement as encouragement and praise for good work changes the behaviour of the learner and creates a positive attitude.*

  Reinforcement can take place verbally through words such as 'good', 'very good' and 'excellent' or non-verbally through a friendly smile, an approving nod or a meaningful wink.

- *Questioning - Questioning, as a special technique, should be one of the tools of trade of every teacher, and yet it is one aspect of teaching that is often neglected, possibly because it is so obvious.*

Training programmes to promote verbal interaction in the classroom will also concentrate on questioning. The following sub-criteria can be applied to evaluate such presentations:

* Unambiguous formulation - The learner must know exactly what is being asked because ambiguity can lead to obscurity.

* Tempo of questioning - Time for thought must be allowed for more difficult questions. Teachers sometimes undermine the objective of questioning by prematurely answering their own questions.

* Distribution of questions - Unless all the learners are involved in answering questions, some may feel redundant.

* Types of questions - Convergent and divergent questions should be alternated.

* Prompting techniques - These should be used to help learners find the right answer or to formulate answers correctly. Correct answers are reinforcing and will provide further motivation to learners.

* Discipline in questioning - Learners should be disciplined to answer questions only when they are asked to do so. If this is not done, the most important object of questioning, namely to actualise thought, may be nullified.

OUTCOME CRITERIA

The criteria mentioned below will be applied to test the results/effects of a programme after it has been broadcast. Feedback is necessary at this stage to determine whether the programme will be released for future use, and if other similar programmes will be considered. Information collected in this manner can also be of great value in the production of other training programmes.

- Completeness and suitability of the content - This 'completeness' criterion is applied to determine whether the programme has provided enough information to clarify a certain topic, idea or concept. It is possible in making a programme on Computer Science, for example, to assume wrongly that the target group has certain basic knowledge: the programmes would then be lacking in content. The criterion may also indicate, however, that because too much information has been included, the learners are confused and bored.

  The 'suitability' criterion can be applied to determine whether the target group has found the programme relevant or appropriate to their needs. Evaluation may show that the standard of a programme for a certain class is too high or too low: it may fascinate and stimulate the imagination of boys, yet may be regarded as a complete failure by girls of the same age-group.

- Realization of objectives - Ultimately, programme compilers want to determine whether the training programme has been effective. This brings the whole question of learning realization into focus. A distinction must be made, however, in learning. "In its simplest form, it is a distinction between process and outcome" (Gooler, 1977:84). Although they are very difficult to determine, the outcomes of learning should be judged individually and collectively, and in the long-term.

Cost-effectiveness - The cost-effectiveness of a programme can be determined in the first place by comparing the costs of similar programmes. Secondly, costs should be related to programme efficacy. The decision as to whether a programme should be incorporated in a training course is determined, to a great extent, by its cost-effectiveness. The cost-benefit parameter is strongly supported by economists although its applicability to education is highly questionable (Lewy, 1985:22).

Conclusion

The objectives of the programme will determine which criteria to use in the evaluation of teacher training programmes. Practice-based criteria may be useful for evaluating programmes dealing with the behaviour of the teacher in the classroom, whereas programmes dealing with the teaching of a language will be evaluated mainly by theory-based criteria. Criteria for evaluating the production of programmes will be applied to most programmes.

REFERENCES

Bates, A 1984. Broadcasting in Education. London, Constable and Co.

Brown, G 1975. Micro-Teaching a Programme of Teaching Skills. London, Methuen and Co. Limited.

Brusling, C 1974. A Concept in Development. Stockholm, Sweden, Almquist and Wiksell International.

Cawood, J 1982. Die moontlikheid en noodsaaklikheid van voortgesette onderwys vir volwassenes (The potential of and need for continuing education for adults). The Role played by the University in Non-formal Continuing, Community and Adult Education. Workshop 3-4 May 1982. Stellenbosch.

Coldevin, G O 1981. Experimental research in television production techniques and presentation strategies: Current Directions. Teaching Strategies.

Cropley, A J 1977. Lifelong Education. A Psychological Analysis. Oxford, Pergamon Press.

Ebel, L 1974. What do tests test? Symposium on Evaluation in Education, Johannesburg College of Education, Johannesburg.

Fauquet, M 1971. Orientation on research on the use of closed-circuit television for teacher training. Educational Media International.

Gooler, D D 1977. Criteria for evaluating the success of non-traditional postsecondary education programs. Journal of Higher Education, Vol. XLVIII, No. 1, January/February 1977.

Hawkridge, D and Robinson, J 1982. Organising, Educational Broadcasting. London, Croom Helm.

Howe, J A (ed.) 1983. Learning from Television. London, Academic Press.

HSRC Teaching Research Programme 1984. Die benutting van die radio en televisie in onderwys en opleiding (The use of the radio and television in Education and Training). Report of the Main Committee of the HSRC Teaching Research Programme. Pretoria, Human Sciences Research Council.

Kowalski, T J 1984. Research and assumptions in adult education: implications for teacher preparation. Journal of Teacher Education, Vol. XXXV, No. 3, May/June 1984.

Knowles, S 1980. The Modern Practice of Adult Education. From Pedagogy to Andragogy. Chicago, Follett Publishing Company.

Lee, R H 1982. Design of a general structure for the provision of non-formal education. The Role of the University in Non-formal Continuing, Community and Adult Education. Workshop, 3-4 May 1982. Stellenbosch.

Lewy, A 1985. Programme evaluation features invariant across entities being evaluated. Educational Programmes RTV1. Vol. 1, No. 1, 1985.

Liebenberg, C R 1978. The Implementation of Television in Teacher Training in the RSA. South African Human Sciences Research Council, Pretoria.

McDaniel, T R 1982. Teaching television literacy to teachers. Educational Technology, Vol. 22, No. 8.

Miron, M and Segal, E 1978. 'The good university teacher' as perceived by the students. Higher Education, Vol. 7, 1978.

Mobbs, J B 1984. Evaluering van Onderwysmedia vir die Onderwys en Onderwysersopleiding (Evaluation of Teaching Media for Teaching and Teacher Training). Unpublished M.Ed. thesis. University of Port Elizabeth.

Ornstein, A C 1976. Can we define a good teacher?. Peabody Journal of Education, April, 1976.

Reed, K 1976. The role of videotape recordings in the training of teachers. International Journal for Instructional Media, (3)(2), 1976.

Schramm, W (ed.) 1972. Quality in Instructional Television. An East-West Center Book. Honolulu, The University Press of Hawaii.

Schramm, W 1977. Big Media Little Media. Tools and Technologies for Instruction. Beverley Hills, SAGE Publications.

Shively, J E 1975. Evaluation of the technical quality of educational television programs. Educational Technology, Vol. 15, No. 11.

Slabbert, B R 1982. Die Aanbiedingstrategieë van Voortgesette Onderwys aan Universiteite (The Presentation Strategies of Continuing Education at Universities). The role of the University in Non-formal Continuing Community and Adult Education. Workshop, Stellenbosch, 3-4 May, 1982.

Van der Stoep, F and Louw, W J 1976. Inleiding tot die Didaktiese Pedagogiek (Introduction to didactic pedagogy). Pretoria, Academica.

Wessels, H J 1982. Vakdidaktiek van Algemene Wetenskap vir die primêre skool (Subject didactics for General Science for primary school). Pretoria, De Jager-HAUM Publishers.

EVALUATING EDUCATIONAL TELEVISION PROGRAMMES PRODUCED FOR PRE-SCHOOL CHILDREN

E Nel and P de Vries

For Wartella (1979), the pre-school child's life is centred around its family, possibly supplemented by contacts with teachers and peers in day-care centres or nursery schools. At about six years of age its view of life is broadened to encompass a new developmental influence: formal schooling.

This view of a pre-schooler is challenged by Lesser (1974) who points out that most young children do not have the advantage of a pre-school education. Pre-schoolers who have had little or no wider educational stimulus prior to formal schooling could start school at a disadvantage vis-a-vis children who have had educational stimuli.

ETV was introduced in the USA primarily to meet the educational needs of pre-schoolers in lower income groups who were not receiving educational stimulus. Because nearly every home boasted a TV set, the television network could provide a surrogate teacher (Rapaczyniski et al, 1982). The teaching phase of television for children in the USA began with commercial broadcasts in 1940. Puppet shows and short programmes for children were broadcast along with cooking demonstrations, sport, news and films. The development of further educational programmes was delayed by World War II, and the subsequent freeze on licensing by the Federal Communications Commission (FCC). However, two major breakthroughs occurred in 1952:

- the FCC's registration of 242 television channels for education
- the establishment of the Educational Radio and Television Centre (In 1958 the Centre became known as the National Educational Television - NET).

Television has only been part of every-day life in South Africa for about ten years. Human Sciences Research Council (HSRC) figures show that children spend an estimated average of 19 hours viewing television per week ("Die Burger", 1955). This provides an opportunity for producers and directors to orientate ETV towards education, and to prepare pre-schoolers for formal schooling.

The problem is how to ensure that the type of programme broadcast is appropriate to the pre-school child's stage of development. This is where evaluation can play a major role in setting criteria for ETV-programmes for pre-school children, and determining whether these criteria have been met. The following description of the development phase of the pre-schooler can be used to develop criteria for evaluating ETV-programmes for children in this age group. An evaluation of this nature - particularly if it is done formatively - can assist the production team to make ETV-programmes relevant to the pre-schooler's phase of development. This chapter also provides a short description of the goals that ETV can serve. Clearly, the evaluator will need to marry the two elements: the pre-schooler's development stage with the goals that the ETV-programme hopes to achieve to assist the production team to orientate the programme to the appropriate end product.

Profile of the Pre-school Child

AGE

From ages one to four a major increase in attention to television takes place. Informal observation has shown that children younger than 30 months of age do not monitor the screen systematically, but have their attention 'captured' for short periods of time. They appear more interested in playing with toys and interacting with their mothers than watching television.

Other children appear to 'watch' television more deliberately: they sit orientated towards the set, often playing with toys, and glance up at the screen frequently. Older children look at the screen for up to seven minutes without glancing away, whereas the longest viewing period of the younger children is about 60 seconds (Anderson and Levin, 1976).

Tower et al (1979) have suggested that children tend initially to encode visually, and if permitted ample time, will label images. As the child grows older, it is able to encode and store verbally as well. When the child is denied rehearsal and labelling opportunities learning is impaired (age is not a significant factor). Learning appears to be a function of experience as well as ability.

GENDER

The gender of the child has a strong influence on its viewing habits, but this variable is also influenced by the child's IQ. Tower et al (1979), tested the performace of pre-school children on factual inferential recall and recognition questions. They found that girls with high IQ scores performed significantly better than did boys on all the questions, except for the recognition questions.

This finding indicates that young girls generally process verbal information quicker and more effectively than do young boys. The superior recognition memory of the boys suggests that imagery may be the preferred mode of presentation for small boys. In essence gender differences in terms of learning are not clear or significant: IQ plays a much more important role in determining individual differences.

Something of interest which concerns the quality and quantity of attention much more than the gender of the pre-schooler, is the gender of the characters. Attention is lower for male characters and enhanced for females (Anderson and Levin, 1976). They suggest that this phenomenon was due to young children finding adult men inherently less interesting than women - a notion supported by work on social reinforcement.

COGNITIVE DEVELOPMENT

As children investigate the world around them, they acquire cognitive skills for gathering and organizing new knowledge. These information handling skills can be observed in many situations where children pay attention, recall and solve problems strategically.

Attention (See paper by A Parekh for more extensive coverage)

The allocation of attention is important. When presented with an array of pictures and asked to name the items, older children between three and eleven years of age performed progressively better and faster than younger children. They scanned the pictures more exhaustively, focussed more on salient information and made fewer errors (Lorch et al, 1979; Wolman, 1982).

Eye-movement recordings have shown that four-year olds did not examine all the features of the pictures systematically, whereas six to nine year olds did. Before the age of five, the more salient aspects of the pictures capture children's attention. Poorer performance of young children in directing attention and visual scanning is not simply an inability to understand the stimuli, but appears to be a failure to impose a strategic organization on the task (Wolman, 1982).

It follows that ETV must provide a sound structure in its presentation of material to facilitate learning. The learning material must thus not only be understandable but must also be presented strategically (Blake, 1979).

Attention is furthermore enhanced by:

- Presence of women
- Presence of children (i.e. viewing in groups)
- Eye contact

- Puppets
- Peculiar voices
- Animation
- Movement
- Lively music
- Rhyming
- Repetition and alliteration
- Auditory change (Anderson and Levin, 1976).

Both theoretical and empirical data indicate that a child learns best that to which she pays attention. Attracting and maintaining attention should thus be the objective of ETV producers (Anderson and Levin, 1976).

Memory Skills

Although the ability to remember past experience is evident from early infancy onward, some aspects of memory reveal more striking developmental changes than others. Recognition memory, for example, is very accurate throughout childhood. General tasks which are not necessarily meaningful to the child and which require no special recognition strategies reveal slight performance changes during childhood (Wolman, 1982).

Wolman (1982) cited a study where pictures were shown to four, seven and eleven-year olds under two instructional conditions: 'Look at the pictures' or 'Remember the pictures'. The four year olds responded in the same way to both instructions, whereas older children employed study strategies when trying to remember. Other evidence has indicated that the problem of retention may be one of novelty. Information that is provided for novelty sake, that requires artificial strategies or that has an unclear goal, will not enhance retention. Pre-schoolers seem to react to a certain 'naturalness' of the material (Wolman, 1982). Islomina (1982) has argued that the purposefulness of the child's task must be natural and motivating and that children "must not only know what remembering is by itself, but must also be able to see it as an end result, an objective to which activity must be directed" (Wolman, 1982:336).

The educational programmer must bear this in mind. The goal must be there for the child to achieve and she must have no doubts about the value and sense of learning the presented material. The moment the material becomes 'irrelevant' or 'boring' to him, his attention shifts and no learning takes place.

Language and Thought

Pre-schoolers often have difficulty distinguishing words that signify relationship (e.g. quantity, size, direction). Part of the difficulty seems to arise from the imprecise way young children acquire these terms from adults. It is probable that formal instruction on television can teach children to use words more accurately. This training would then focus on general education in the structure and function of language.

Educational television can also teach children to use language creatively. In this manner the use of relational terms ('Mike has more than Mary'), differentiated terms (different terms for different dimensions) and co-ordinated terms ('This pencil is longer but thinner, the other is short but thick') can be taught.

Lesser (1977) cites evidence that children have difficulty separating words in a spoken sentence and sounds in a spoken word. This can be rectified by instructional programmes requiring children to distinguish each word or sound. Since sentence-word and sound-word analysis are preconditions for literacy training, ETV should concentrate on this technique.

Another important consideration in training children in language usage is to teach them to rely on verbal reasoning, even when it is in conflict with an observed event. This is best described by an example:

> *The pouring of water to and from a narrow and a wider container behind a screen - i.e. hidden from the eyes - should force children to rely exclusively on verbal reasoning.*

> *They may conclude that the same water is used in each action, and may not become confused by the difference in containers. Such a presentation is easily done on television (Lesser, 1977).*

However, cognition, language, perception and memory do not exist in isolation. Each is a different aspect of the overall process of development.

A review of cognitive development emphasises that 'learning how to learn' is a real and important aspect of a young child's development. Pre-schoolers require preparation before exposure to school skills. If ETV could instill sufficient reasoning skills in children, exposure to skills would be readily, flexibly and creatively acquired (Salomon, 1977).

SOCIAL-EMOTIONAL DEVELOPMENT

Most countries of the world maintain some form of social stratification or social class system. The strata within a social system may vary, but the classification of rich and poor, upper and lower class is fundamental to the social structure.

A widely available and heavily used medium of communication can present broad representations of different social classes: these representations are important in the social learning of children. As a social communicator, television can reinforce selected concepts about social status (Adler et al, Rossiter and Ward, 1980; Plamer and Dorr, 1980).

Liebert et al (1973) claim that television communicates information about social structure and shapes attitudes about ourselves, others and the world.

The child can also be orientated to his place and role among others and his relationships with them. ETV teaches social conduct through modelling and reinforcement. In this manner social and personal data are conveyed to the child (Glickman, 1979). Friction can also be educational but on a less intellectual, and a more emotional or personal level. The child can thus be acculturalized and educated according the Western tradition and institutions (Nel, 1979).

The child can be made aware of moods and feelings within himself. He can learn to recognise these emotions in others. He can also develop an emotional vocabulary (Cater and Adler, 1975).

The main socio-emotional functions of ETV could comprise the 'civilizing' of the pre-schooler by confronting him with social beliefs and practices. He can learn about his relationship with other people and how to cope with them (Coates et al, 1976; Comstock et al, 1978).

SENSE OF THE SELF

The child with or without pre-school educational experiences, goes to school with a certain self-image. This has been influenced by interaction with family members and others in his social environment. He has certain expectations of school which affect his total school education. These expectations also influence his self-concepts, thus creating a vicious circle.

Successful educational television may eventually result in the child developing a feeling of worth and capability - two concepts integral to her self-concept. The child who has mastered some skills from television viewing is likely to come to school with a certain composure and confidence which a non-viewer would normally lack. If ETV helps a child view himself positively then it is indeed successful. ETV could be used to expose virtually all children to some form of self- or ego-boost.

The child's peer group and social interaction are also important factors determining her sense of self. Successful relationships, friends and the experience of feeling liked or popular help the child view herself positively. Unfortunately, social incidents, quarrels and social isolation leave emotional scars which are not easily healed.

The last - and most important - premise for cultivating self-esteem is the family. Should the child grow up in a climate of empathy, understanding and acceptance, she will come to terms with herself and those around her (Van Wyk, 1983). These are the corner stones for building self-esteem which can never be stressed enough (Van Wyk, 1983). A child growing up in a climate devoid of empathy, understanding and empathy will tend to be troubled, uncertain and feel unimportant.

Self-esteem can be brought about by the family and the social environment. In both instances ETV has a major role to play. When compiling and producing an educational programme the question which must always be kept in mind is: 'Will this help the child view herself in a positive manner?'. It is important to stress the potential and capabilities of the pre-schooler. No educational activity can be undertaken without assessing and honouring the childs needs and potential. Learning can take place best if the activity is directed at the viewer, and attuned to her needs.

Learning and the Pre - Schooler

Researchers have discovered wide variations in viewing styles among children. However, the different styles do not seem to relate closely to how much a child learns. This work has increased interest in viewing styles. Lesser (1974) has categorised the viewing styles as follows:

* Children seem to watch while they simultaneously keep track of other interesting events around them.

* 'Zombie viewing' - which may reflect either intense concentration or stupor.

* Over-active physical and verbal participation in the televised action.

No matter what the pattern, it is hoped that the child does learn something from his television viewing. Research has provided us with certain clear-cut facilitators of learning for implementation by ETV-programmers.

FORMAT AND CONTENT

Learning is triggered not only by content but also by format. Format cues can provide prompting devices for learning. They can also be called 'conventions' of the medium and include:

- swelling music
- dramatic pauses
- fade-outs, zoom-ins and freeze-frames

MODELLING

Children imitate televised characters. This leads to concern about agression and violence being modelled. The more the child identifies with the character, the better model the character is and more likely the child is to copy his behaviour (Sproull, 1973).

Learning in this manner need not only be direct, but can also be indirect or implicit. In this manner altruism, kindness, courage and tolerance can be communicated through televised actions (Duck and Baggaley, 1975). The importance of modelling is best conveyed by an exclamation from a black child while watching 'Sesame Street'.

> *"Look! He looks like me and he knows the answers!" .... (Lesser, 1974:86).*

NARROW FOCUSING

Pre-schoolers have difficulty discriminating between what is essential or relevant and what is incidental or irrelevant. Extraneous material - visual and auditory - must be reduced, or better still, eliminated. The rule of 'the simpler the better' must be applied. Deciding what is relevant is a problem, because material which someone might describe as irrelevant might be seen as very relevant by another. Television can provide the young child with a precise focus on central content, carefully eliminating irrelevancies and distractions.

CROSS-MODAL REINFORCEMENT

Researchers have spent hours arguing whether words or pictures work better.

They have, however, not reached any agreement as to whether the two modes of learning should be separated or used in conjunction with each other. Lesser (1974) favours a learning scheme where one modality is used to support another, instead of cancelling out or interfering with the other.

DIRECT AND INDIRECT TEACHING

Direct teaching involves telling and showing a message and then repeating it before signing off. Only a small part of what is learned is taught directly (Wilkens, 1982).

Most television messages are indirect and implicit. Social development especially is facilitated by this means, of which modelling is the most common method.

'SHOWING THE WORLD'

Lesser (1974) uses the term 'showing the world' to describe television's capacity to teach children about the world. Knowledge of people, events and new ideas is conveyed. Yet, who will make the final decision on what should or should not be shown?

Should only the nice and sweet things of life be brought to their living rooms or should programmes be more realistic?

The producers of 'Sesame Street' chose a 'sweet' show - and are still defending their stance. Lesser (1974), however, concluded:

> *The harsh realities of a child's world surely exist. We lose credibility with the child if we ignore them entirely, and no effort to teach can afford this. But, even at the risk of sugar coating these realities, perhaps we can suggest the vision of better things (Lesser, 1974:95).*

'PREACHING'

'Preaching' to children sometimes works, but the literature is quite clear that 'preaching' on television simply does not work. The audience chooses freely what it wishes to view and has the right to its own conclusions about the virtue and morality of a programme. Care should be taken to refrain from 'preaching'.

TRIVIALIZING

For many of the same reasons, a 'talking to children' tone of voice must be guarded against. This takes the form of a sing-song voice, raising a few tones and lisping a little and is belittling. Trivializing and simplifying are not synonymous, and children probably know the difference well.

Goals of ETV

Goals depend on the taste or wishes of the producer or production team and can thus vary. At first the 'Sesame Street' team placed a high premium on preparing disadvantaged children for school. Emphasis was on reading and general pupil skills (Rankin and Roberts, 1981), but later their goals changed.

Goals can be divided into four categories:

SYMBOLIC REPRESENTATION

This category has letters, numbers and geometrical figures as its content. Its objective is to enhance understanding and the proper use of the content. This implies knowledge of labels and the relation of characters to one another, with eventual school readiness in mind (Van Zyl, 1974).

COGNITIVE PROCESSES

Cognitive goals include the enhancement of:

Perceptual discrimination:

- Body perceptions
- Visual discrimination
- Auditory discrimination

Relational concepts:

- Size relationships
- Positional relationships
- Distance relationships
- Amount or number relationships
- Temporal relationships
- Auditory relationships

Classification:

- Size
- Form
- Function
- Class

Ordering:

- Size
- Temporal sequence

Reasoning and Problem Solving:

- Inferences and causality
- Explanations and solutions
- Attitudes (persistence, impulse control)

(Lesser, 1974; Van Zyl, 1974)

PHYSICAL ENVIRONMENT

The child's conception of the physical world should include general information about natural phenomena, certain processes and certain interdependencies.

The physical world around him:

- Land, sky, water
- City and country
- Plants and animals
- Natural processes and cycles (reproduction, growth and development; weather and seasons)
- Machines
- Buildings

SOCIAL ENVIRONMENT

Social units:

- Self
- Roles
- Social groups and institutions

Social interactions:

- Differences
- Co-operation
- Rules (Lesser, 1974).

ETV thus strives to educate and inform children on one or more of these topics.

Summary

The following is a list derived from Lesser (1974) of the procedures employed by 'Sesame Street' to induce learning:

- Learning may occur through modelling or observation. This is also called 'indirect teaching' or the use of role modelling and situation-shaping to elicit imitation of behaviour and to change attitudes (Moir, 1967; Polsky, 1974).
- Learning may occur through direct teaching by showing or telling what you intend to teach; through teaching and through telling what was taught (Hilliard, 1967; Lesser, 1977).
- Learning may occur if children are motivated. Lesser (1977) believes that 'children learn best what they want to learn'.
- A principle pre-requisite for learning is attracting, directing and sustaining the child's attention. Attending is thus equated with learning (Lesser, 1974; Lesser, 1977; Moody, 1980).

The above-mentioned procedures are the background against which learning takes place. Particular facilitators of learning are also employed in the form of techniques such as those mentioned in the section on 'Learning and the Pre-schooler'.

In answering the question: 'Why do children watch TV?', and obvious answer springs to mind: Simply because it is there!

Research has shown that children use TV both for information gathering and as an escape, as a social tool and in order to withdraw from real life (Blake, 1979). Schramm et al (1961) have found that television meets mostly the fantasy needs of children. Television invites them to leave the problems of the real world and offers instead relaxation, pleasure and wish-fulfilment.

Evaluation of, and research into, educational television is of utmost importance. The process of education must be scrutinized continually and care must be taken to ensure that the particular needs and potential of the target-viewer are never forgotten.

REFERENCES

Adler, R P, Lesser, G S, Meringoff, L K, Robertson, T S, Rossiter, J R and Ward, S 1980. The Effects of Television Advertising on Children. Toronto, D.C. Heath and Co.

Anderson, J 1982. Television codes and audience response. Media in Education and Development, 24-29.

Baggaley, J and Smith, K 1982. Formative research in rural education. Media in Education and Development, 15(4), 173-176.

Blake, F D 1979. The educational significance of television. Educational Journal, 80(3), 13-23.

"Burger, Die". TV1 - Violence more than in the USA. 25 June 1985.

Burke, R C 1971. Instructional Television: Bold New Venture. London, Indiana University Press.

Cater, D and Adler, R 1975. Television as a Social Force: New Approaches to TV Criticism. New York, Praeger Publishers.

Coates, B, Pusser, H E and Goodman, I 1976. The influence of 'Sesame Street' and 'Mister Rogers' Neighbourhood' on children's social behaviour. Child Development, 47(4), 138-144.

Comstock, G, Chaffee, S, Katzman, N, McCombs, M and Roberts, D 1978. Television and Human Behaviour. New York, Columbia University Press.

Duck, S W and Baggaley, J 1975. Audience reaction and its effect on perceived expertise. Communication Research, 2, 79-85.

Field, S 1974. Professional Broadcast Writer's Handbook. Blue Ridge Summit, Tab Books.

Glickman, L J 1979. Footsteps toward understanding children. American Education, 15(1), 12-17.

Hilliard, R L 1967. Writing for Television and Radio. New York, Hastings House Publishers.

Human Sciences Research Council (HSRC) 1984. The Use of Radio and Television in Education and Training. Newsletter 4, Bureau for Public Relations.

Human Sciences Research Council (HSRC) 1984. Children's Reactions in front of TV. Newsletter 151, Bureau for Public Relations.

Joffe, B H 1981. Television as an educational institution. Media and Methods, May/June, 12.

Lesser, G S 1974. Children and television: Lessons from 'Sesame Street'. New York, Vintage Books.

Lesser, H 1977. Television and the Pre-School Child: A Psychological Theory of Instruction and Curriculum Development. New York, Academic Press.

Liebert, R M, Neale, J M and Davidson, E S 1973. The Early Window: Effects of Television on Children and Youth. New York, Pergamon Press Inc.

Lorch, E P, Anderson, D R and Levin, S R 1979. The relationship of visual attention to children's comprehension of television. Child Development, 50(3), 722-727.

Mander, J 1978. Four Arguments for the Elimination of Television. New York, Morrow Quill Paperbacks.

Moir, G 1967. Teaching and Television: ETV Explained. New York, Pergamon Press Inc.

Moody, K 1980. Growing up on Television: the TV Effect. New York, Times Books.

Nel, E M 1979. Television and children. Social Work, 15(2), 69-74.

Nel, E M 1981. The Child and His Leisure Time: Television, Radio and Films. Paper read at Autumn School, University of Stellenbosch Medical School, Tygerberg.

Palmer, E L and Dorr, A 1980. Children and the Faces of Television: Teaching Violence and Selling. New York, Academic Press.

Pierce, C M (ed.) 1978. Television and Education. California, Sage Publications.

Polsky, R M 1974. Getting to 'Sesame Street': Origins of the Children's Television Workshop. New York, Praeger Publishers.

Rankin, P M and Roberts, C W 1981. Television and teaching. The Reading Teacher, 35(1), 30-32.

Rapaczynski, W, Singer, D G and Singer, J L 1982. Teaching television: A curriculum for young children. Journal of Communication, 32(2), 46-55.

Rushing, B 1982. Cognitive style and videotapes. NSPI Journal, October, 36-39.

Salomon, G 1976. A cognitive approach to media. Educational Technology, 16(5), 25-28.

Salomon, G 1977. Effects of encouraging Israeli mothers to co-observe 'Sesame Street' with their five-year-olds. Child Development, 48(3), 1146-1151.

Schleicher, K 1971. The necessity of an anthropo-ecological dimension in education - possibilities for its realisation. International Review of Education, 21(2), 231-260.

Schramm, W, Lyle, J and Parker, E B 1961. Television in the Lives of Our Children. Stanford, Stanford University Press.

Sproull, N 1973. Visual attention, modelling behaviours and other verbal and non-verbal meta-communication of prekindergarten children viewing 'Sesame Street'. American Educational Research Journal, 10(2), 101-114.

Tower, R B, Singer, D G, Singer, J L and Biggs, A 1979. Differential effects of television programming on pre-schoolers' cognition, imagination and social play. American Journal of Orthopsychiatry, 49(2), 265-281.

Van Wyk, J D 1983. The Relationship between Parent and Child. Cape Town, Human en Rousseau.

Van Zyl, P J 1974. Objectives of TV for Pre-Schoolers. Onderwysblad, 81(846), 7-11.

Wartella, E (ed.) 1979. Children Communicating: Media and Development of Thought, Speech and Understanding. California, Sage Publications.

Wilkens, J A 1982. Breaking the TV Habit. New York, Charles Scribners and Sons.

Withey, S B and Abeles, R P 1980. Television and Social Behaviour: Beyond Violence and Children. New York, Lawrence Erlbaum Associates.

Wolman, B B (ed.) 1982. Handbook of Developmental Psychology. New Jersey, Prentice-Hall Inc.

Part Three: Investigations

THE EFFECT OF FORMAL FEATURES ON THE ATTENTION GIVEN TO AN ETV-PROGRAMME BY SOUTH AFRICAN INDIAN PRE-SCHOOL CHILDREN

A Parekh

Introduction

McLuhan (1964) proposed that the dominant mass media convey the social image and set the agendas that shape a generation's weltanschauung. The codes and formats of electronic media in particular will change fundamentally ways of thinking and communicating among those children who spend most of their time processing its messages. A mass of research into the effects of formal features of television has resulted from McLuhan's statement.

Formal features of importance are camera techniques (cuts, pans, zooms and special effects), auditory properties (dialogue, laughter, track and music), rates of scene and character change, and action of characters and objects. Formal features result largely from production and editing decisions: they can be used with a wide range of content (Huston et al, 1982). Content refers to the thematic material that is special to the particular drama, documentary, or educational message in the production (Huston et al, 1982).

Form and content can be defined independently, yet there are familiar conventions for using certain forms with particular categories of content. Masculine and feminine toys are presented with distinctly different formal features. Soft, fuzzy images and background music on television are associated with feminine products, whereas masculine products are advertised with high action, quick cuts, loud music and sound effects (Welch et al, 1979).

Huston et al (1981) demonstrated that in commercial programmes made for children, the formal feature, animation, is associated closely with violent content. Other features which accompany violence are high rates of physical action, loud music, sound effects and visual special effects. There is a close association between form and content in programmes for young children. Huston and her colleagues (1982) showed that the younger the child, the more likely he is to attend to these dazzling techniques of television, regardless of programme content.

By contrast, older children watch programmes for extended periods even when such formal features are not heavily exploited. Levin and Anderson (1976) and Calvert et al (1982) showed that formal features such as animation enhance children's attention to television. Campbell (1983) studied nutritional announcements containing identical visual and verbal content. Two types of formal features were selected to represent forms typically used in adult and children's programmes. The adult versions were made from live footage with an adult male narrator and staid background music, and the children's versions were animated with peculiar character voices and lively whimsical music. It was found that pre-school children attended more to the child - than to the adult versions of the content.

A limitation of the studies cited above is a failure to examine content independently of form. A study which attempted to overcome this limitation was conducted by Fotts et al (1982). They designed a study to investigate the separate effects of violent content and formal features on children's attention and social behaviour, using a broader sample of programmes than was used in previous studies. It was found that rapid action and animation clearly enhanced children's attention, but that violent content had little effect on attention.

Besides these intra-programme factors, there are also extra-programme attractions - such as toys, friends, etc. that play a role in the amount of time the child actually spends watching TV programmes. Lorch et al (1979) found that without toys or other activities to provide alternatives to watching television, visual attention can be very high. However, Anderson et al (1981) found that despite the presence of certain formal features, children lost interest when the programme was made incomprehensible by scrambling the auditory track, or using a totally foreign language. Many educational television producers believe they need to rely on content in order to gain the child's attention and influence his behaviour, rather than include in their programmes those formal features which enhance children's attention and influence their behaviour.

Theoretical Framework

Earlier research on children's television viewing has been influenced heavily by the cognitive and social-learning models. These theories differ not in the conclusions they reach, but in the areas they examine: cognitive theory concentrates on the process by which children assimilate knowledge; social-learning theory is concerned with the effects of television on children. The premises of the latter theory have been brought into serious doubt by the adherents of arousal theory. In order to understand the influence of these theories on research, a brief discussion of each is necessary.

COGNITIVE THEORY:

Children's information gathering and processing is a two-fold process: there is an initial attraction to perceptually salient features, and this is followed by a focus on features that are informative in the structure and content of the televised material. Attention is viewed as the first stage of information processing or the funnel through which all chains of influence must pass.

One principle used to explain attention to television which may be described as continuous attraction, is called "attentional inertia" (Anderson et al, 1979). Stated simply, the principle is that the longer people look at the television screen, the greater is the probability that they will continue to look. Attentional inertia appears in children as young as a year old, as well as in older children and adults. The converse, not looking at television, is also related to inertia: the longer children direct their attention elsewhere, the less likely they are to turn their attention to the screen.

Whilst attention may thus be regarded as a function of the viewing time-span, recent research has shown that it is determined by a second factor: children are also selective in what they attend to on television. Selective attention is a principle that has been used to explain how children, and perhaps adults, manage to avoid being bombarded and overwhelmed by the vast array of information that is presented on television. In other words, although a child may sit in front of a television set following the whole programme, his attention is not on the programme throughout this period.

The perceptual salience of a stimulus which is a determinant of attention is important in determining the attention of children who are not familiar with television. This theory is based on research findings in child development, indicating that as children get older they move from "exploration" to "search" to obtain information (Wright and Vlietstra, 1975). In the "exploration" stage, children respond to the most salient features of their environments, and they respond in short, discontinuous and impulsive bursts to whatever is perceptually dominant at any time. Applied to television viewing the hypothesis is that, for the youngest and least experienced viewers, television viewing consists of absorbing the perceptually salient parts of a programme as entertainment in its own right. The very young child is essentially a passive consumer of audio-visual thrills.

Attention has also been found to be linked to the comprehensibility of the programmes. Formal features, like animation or children's voices, are signals to children that the programme is designed for them. From their experience with television, children learn to interpret these signals (Lorch et al, 1979).

The more recent trend in thinking among psychologists and researchers is that children play an active and selective role in how they approach television as a medium. The stereotype of television viewing as a passive, intellectually lazy occupation is rejected. Recent research on cognitive processing has advanced our knowledge of this issue by showing that young children are actively processing the content of television. However, on account of its subject matter, this theory does not address the issue of the behavioural effects of the medium. The theoretical model that has been used most extensively to account for such effects is the model articulated by Bandura (1977): observational learning theory.

OBSERVATIONAL LEARNING THEORY:

Observational learning theory has formed the basis for much of the research dealing with the effects of television on children. A major premise of the theory is that children learn particular behaviours or categories of behaviour from watching television: when they find themselves in a situation with appropriate cues, they are likely to imitate that behaviour. Although there are many mediating and moderating conditions that influence the likelihood of imitation, a central thesis of this theory is that the child's behaviour is similar to that observed in analogous situations. Researchers such as Huston et al (1982) and Potts et al (1982) suggest that if one accepts the theory that violent content promotes aggressive behaviour then one must also concede that formal features should also lead to behavioural analogies of form, but of a different kind. "For example, high levels of physical action might lead to increased activity levels, or high pace might increase the rate one activity shifts to another. Clearly, the effects of formal features should be different from those of content" (Huston et al, 1982:2).

Although the basic principles of observational learning theory have extensive empirical support (e.g., Bandura et al, 1963; Collins et al, 1974; Eron et al, 1972; Friedrich-Cofer et al, 1979; Huston-Stein and Fox et al, 1981; Newcomb and Collins, 1979; Singer, 1980), many investigators have questioned whether the theory provides a comprehensive explanation of the behavioural effects of television. These critics argue that observations of behaviour of children after they watch violent television, very often reveal that children rarely show exact imitations of what they have seen.

These disagreements have motivated other researchers to propose new theoretical models to account for the effects of television viewing on children's behaviour. A major advance in research on television in the 1970's was the recognition that the medium must be understood in relation to the emotional set of the viewers.

AROUSAL THEORY:

Arousal theory constitutes the other major model used by investigators to explain the effects of television on children's behaviour. General arousal is conceptualized as an activator which increases the intensity of any behaviour; specific behaviours emitted are determined jointly by the individual's pre-dispositions, habits, set and expectations, on the one hand, and by immediate environmental cues on the other (Hull, 1952; Mandler, 1975; Schachter, 1964; Zillman, 1978). Studies have shown that residual arousal from violent, erotic and humourous content of television can be carried over into subsequent situations. TV may activate or intensify a wide range of subjective emotional experiences and behavioural responses. The topography of the behaviour may be quite unrelated to the content of the original source of arousal (Cantor et al, 1975; Mueller and Donnerstein, 1981; Schachter, 1964; Tannenbaum and Zillman, 1975). Hence, the principal difference between arousal theory and observational learning theory is that the former leads to the prediction that both form and content of television can stimulate behaviour that is different from that which occurred in either the form or the content of the programme.

The arousing effects of television form have been investigated by Watt, Krull, and their colleagues (Krull et al, 1977; Watt and Krull, 1974). They proposed that formal features embodying high rates of change, variability, and uncertainty would be cognitively arousing because they would require more intense cognitive processing. Empirical support for their hypothesis can be found in the works of Bryant and Zillman (1979), and Greer et al, (1982). The study by Greer and his associates showed that children were more aggressive after seeing commercial advertisements with high rates than with low rates of action and visual tricks. Neither set contained any violent content. Children in another study tended to be more aggressive after viewing high action than after low action non-violent cartoons (Huston-Stein et al, 1981).

These theoretical models lend support to the notion that the effects of television are due both to form and content.

Learning through imitation of observed actions is a widely held explanation of the influence of programme content. But the emotional and physiological arousal engendered in the viewer can also be invoked to explain television form or code effects.

Present Study

Little or no serious research has been done in South Africa on children's educational programmes. It seems that any first step in this direction would have to include research into the ways and means of improving children's educational television programmes. The aims of the present study were

- to determine to amount of attention pre-school children give to ETV;
- to determine whether pre-school children's attention to television is influenced by certain formal visual and auditory features of educational programmes;
- to determine whether the age and sex of pre-school children affect their viewing; and

- to determine whether Indian pre-schoolers differ markedly from other pre-schoolers regarding the effects of formal features on the attention they give to ETV-programmes.

METHOD

* Subjects: The participants were 11 Indian pre-school children who were selected from a private day-care establishment in Durban and whose medium of communication was English. They were divided into two groups by age (four vs five year olds) and gender.

* Apparatus: The Programme Evaluation Analysis Computer (PEAC) was used to analyse the extent of the children's attention to the television programme. This micro-electronic equipment was developed originally in the studios of the children's television workshop, New York, for formative evaluation research on the world-famous series "Sesame Street" (Mielke and Chen, 1980). The equipment consists of a number of remote, cordless handsets to which an Apple micro-computer is attached.

The television programme used for the research was a 15-minute educational programme intended primarily for the young pre-school child. The programme contained some continuity and key scenes, and major plot resolutions.

The television viewing room was comfortably furnished and designed to make it attractive to children. The children were seated at small tables - about three to four metres in size - in front of the television set. Paper, crayons and a variety of small toys were laid out on the table. The children were told that they could play with any of the materials or watch television, but were asked to remain seated at the table. The purpose of the toys was to provide alternative activities and to make the situation similar in some respects to natural home viewing. The researcher announced that she would be in the next room if she was needed, and then she went into the control room.

The children's visual orientation to the monitor was scored continuously as "off or on" by two observers using hand-held response units. Each observer pressed button A if he saw the child looking at the television set and button B if he observed the child looking away. The one-way mirror through which the observers watched was located directly behind, and just above the television monitor. Videotapes of all the sessions were made.

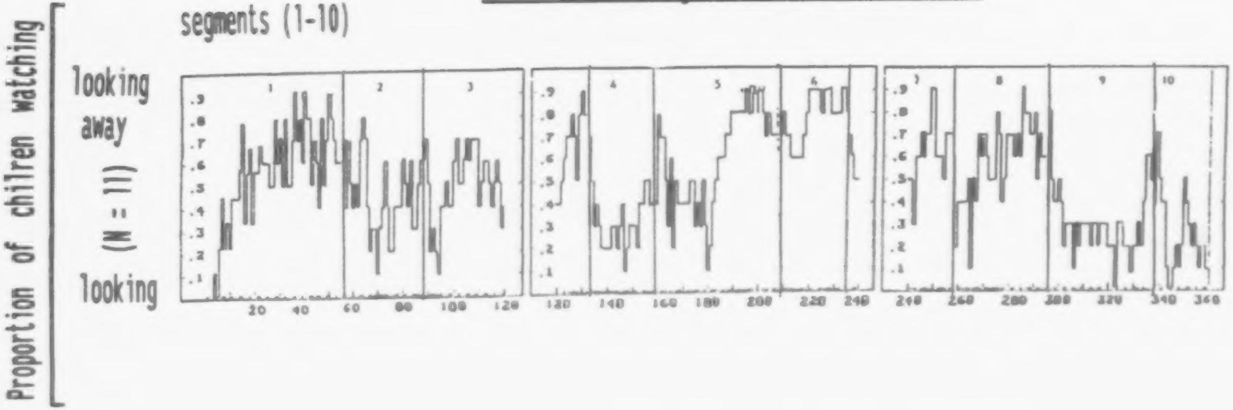
Observer agreement on the same behaviour was 76 percent using the formula: number of agreements x 2, divided by the total number of scores of both observers. The difference between the two observers' scores was not significant (a = 0.05), using the student's T test.

RESULTS

* Amount of attention to programme: The pre-school children watched an average of 50 percent of the programme. There was a general trend for younger children to watch more than the older children (53 percent and 46 percent respectively). These figures are consistent with those found by previous researchers. Anderson and Levin (1976) found that by the age of four, children are watching about 55 percent of the time, even in a playroom with toys, games and other distractions. Similarly, Lorch et al (1979) in a study comparing viewing with and without toys found that five-year-old children given toys to play with during an educational television programme, averaged 44 percent visual attention. There was a minimal difference in the viewing patterns of boys and girls - 48 and 51 percent respectively.

* Effects of formal features: The results collected by the micro-electronic apparatus are presented as mean responses in Figures 1-3. Figure 1 shows that the children's interest fluctuated markedly throughout the programme. The children divided their attention between viewing and not viewing.

Figure 1 : Shifts in children's attention to a 15 minute TV programme - Expressed as a proportion of children watching over 360 time points[1)]



1) 1 TIME POINT = 2,5 SECONDS ⟶

The most plausable reasons for these fluctuations in attention are:

- factors, other than the television programme. The children could have been nervous (despite the fact that every attempt was made to put them at ease), bored or tired. The surroundings also afforded alternatives to watching television. They divided their attention between playing and looking at the television.

  These extra outer-programme factors do not explain why the children did watch on the average 50 percent of the programme, and what factors were responsible for sustaining their attention.

- factors within the television programme ("intra-programme" features), namely, form and content. As the present study failed to examine the effects of form independently of content, one cannot ignore the possible influence of content on attention. However, the question that is being asked is not whether content or form influence children's attention but whether formal features, in any way, attract, sustain, or contribute to the detraction of children's attention. The answer to this question calls for a more detailed analysis of Figure 1 to try and pinpoint those features in the television programme that could possibly determine attention.

In general it was found that children's attention decreased as the length of a segment of the programme increased (for example, bits 1, 3, 5 and 8). On the other hand, children's attention remained constant or in some cases even increased for shorter bits (for example, bits 2, 4, 10).

The children were also attracted to television by specific features of the programme. They more often looked, and continued to look, at such features as female and child characters, puppets, animation and camera activity (such as a zoom-in). However, although puppets enhanced children's attention, animals depressed their attention. Bit 6 which showed a visual of giraffes grazing in their natural environment had very little attentional value for the children. The giraffes were relatively stationary and inactive. Anderson and Levin (1976) found that animals that are relatively stationary and inactive tend to depress attention whereas animals that are active, and in continuous motion, have quite the opposite effect on children's attention.

Lively music was also found to be a powerful elictor of children's attention, though by no means a feature sustaining their continuous attention. Changes in tempo served as important markers that aided comprehension. The children appeared to sense that a change in tempo from fast to slow was a signal to them that a segment was about to come to an end and that a new scene would follow. Towards the end of time points 122-132 the singer began to sing the same song used at the start of the scene at a slow tempo. The children reacted by steadily decreasing their attention.

Figure 2: Shifts in four- and five-year-olds' attention to a 15 minute tv programme - Expressed as a proportion of children watching over 360 time points

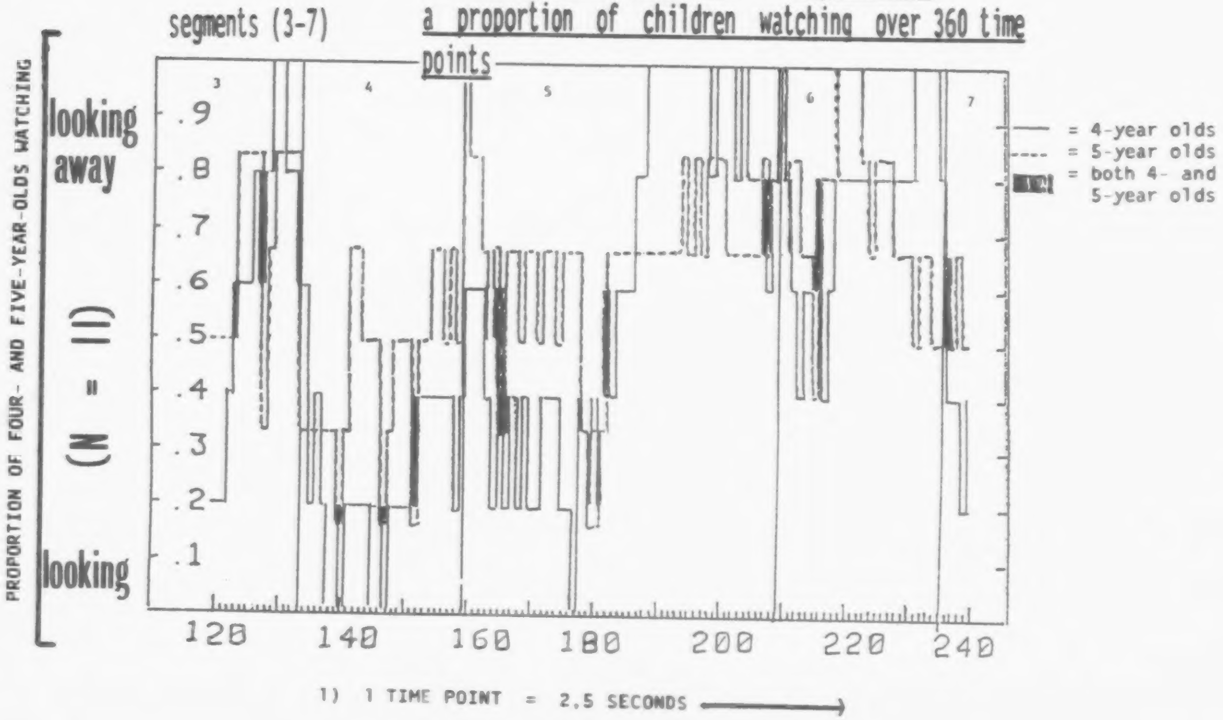
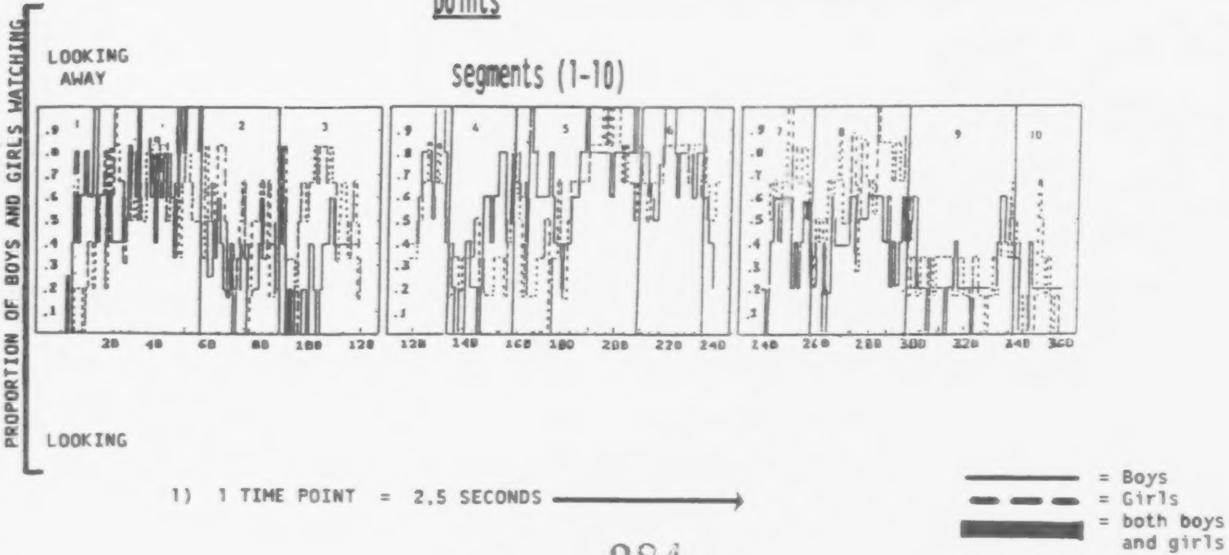


Figure 3: Shifts in boys' and girls' attention to a 15 minute TV programme - Expressed as a proportion of children watching over 360 time points



Other auditory features such as non-human dialogues, child dialogue, and adult dialogue were found to attract and sustain children's attention. When television characters (in this case women) maintained direct eye contact and talked to the children, enlisting their participation in an activity, children were quite attentive.

Presentation of a single letter or a number was found to enhance children's attention. As a method of increasing number and letter recognition, this feature was found to be most effective for pre-school children.

Noise as an attention-grabber was found to depress rather than elevate attention. In most instances, loud non-human dialogue did not attract or sustain children's attention much unless it was coupled with other interesting features. However, humorous programmes that were characterised by non-human dialogue and noise did attract children's attention. Long zooms which involved a close-up presentation of important content did capture children's attention (for example, Bit 9). It proved to be a useful technique for drawing children's visual attention. The attention the children gave to the programme was not influenced noticeably by their age or gender (Figures 2-3).

Programme bits received equal attention. However, there were slightly different preferences within bits. In Bit 4 (Figure 2) the younger children preferred a segment that featured animation. More girls than boys preferred lively music and child characters (time points 95-103); animation (time points 137-147); non-human characters (such as puppets) and humour (time points 240-254).

The present analyses have not eliminated the possible confounding effect of attributes; consequently one is unable to conclude definitively that the presence or absence of an attribute caused attention to be depressed or increased.

The findings of the present study are similar to the findings of similar studies undertaken elsewhere, particularly in the United States and Israel. This similarity deserves close scrutiny by South African producers of educational programmes for children.

Conclusion

The study will need to be replicated with a much larger sample to generate confidence in the generalizability of the findings. In South Africa great stress appears to be laid on cultural differences between the various race groups. Whatever validity that view may hold in other areas, this study suggests that cultural differences do not play a role in the effects of formal features on children's attention. This could serve as a valuable guide to producers who, in attempting to determine the relationship between content and attention, have been conducting the same research four times over - on each occasion using as subjects members of each of the four South African race groups. Producers will save themselves immense costs by designing their programmes for children per se, irrespective of race, as the effects of formal features are likely to be the same.

REFERENCES

Anderson, D R and Levin, S R 1976. Young children: attention to Sesame Street. Child Development, 47, 806-811.

Anderson, D R, Alwitt, L F, Lorch, E P and Levin, S T 1979. Watching children watch television. G Hale and M Lewis (eds.), Attention and the Development of Cognitive Skills. New York, Plenum.

Anderson, D R, Lorch, E P, Field, D E and Sanders, J 1981. The effects of TV programs comprehensibility on pre-school children's visual attention to television. Child Development, 52, 151-157.

Anderson, D R and Field, D E 1983. Children's attention to television: implications for production. M Meyer (ed.. Children and the Formal Features of Television: Approaches and findings of experimental and formative research. Munchen, K G Saur.

Bandura, A, Ross, D and Ross, S A 1963. Imitation of film-mediated aggressive models. Journal of Abnormal and Social Psychology, 66, 3-11.

Bandura, A 1977. Social Learning Theory. Englewood Cliffs, New Jersey, Prentice-Hall.

Bryant, J and Zillmann, D 1979. Effect of intensification of annoyance through unrelated residual excitation on substantially delayed hostile behaviour. Journal of Experimental Social Psychology, 15, 470-480.

Calvert, S, Huston, A, Watkins, B and Wright, J 1982. The relation between selective attention to television forms and children's comprehension of content. Child Development, 53, 601-610.

Campbell, T 1983. Formal cues and content difficulty as determinants of children's cognitive processing of televised educational messages. J C Wright and A C Huston, A Matter of Form. American Psychologist, 38(7), 835-843.

Cantor, J, Zillmann, D and Bryant, J 1975. Enhancement of experienced sexual arousal in response to erotic stimuli through misattribution of unrelated residual arousal. Journal of Personality and Social Psychology, 32, 69-75.

Collins, W A, Berndt, T J and Hess, V L 1974. Observational learning of motives and consequences for television aggression: A development study. Child Development, 45, 799-802.

Eron, L D, Huesmann, L R, Lefkowitz, M M and Walder, L O 1972. Does television violence cause aggression? American Psychologist, 27, 253-263.

Friedrich-Cofer, L K, Huston-Stein, A, Kipnis, D, Susman, E J and Clevitt, A S 1979. Environmental enhancement of prosocial television content: Effects of interpersonal behaviour, imaginative play, and self-regulation in a natural setting. Development Psychology, 15, 637-646.

Greer, D, Potts, R, Wright, J and Huston, A 1982. The effects of television commercial form and commercial placement on children's social behaviour and attention. Child Development, 53, 611-619.

Hull, C 1952. A Behavioural System: An Introduction to Behaviour Theory Concerning the Individual Organism. New Haven, Conn, Yale University Press.

Huston, A, Wright, J, Wartella, E, Rice, M, Watkins, B, Campbell, T and Potts, R 1981. Communicating more than content: Formal features of children's television programmes. Journal of Communication, 31, 32-48.

Huston, A C, Wright, J C and Potts, R 1982. Television forms and children's social behaviour. Paper presented at the annual meeting of the International Communication Association, Boston.

Huston-Stein, A, Fox, S, Greer, D, Watkins, B A and Whitaker, J 1981. The effects of action and violence in television programmes on the social behaviour and imaginative play of the pre-school children. Journal of Genetic Psychology, 38, 183-191.

Krull, R, Watt, J H and Welty, L W 1977. Structure and complexity: Two measures of complexity in television programmes. Communication Research, 4(1), 61-68.

Levin, S R and Anderson, D R 1976. The development of attention. Journal of Communication, 26(2), 126-135.

Lorch, E P, Anderson, D R and Levin, S R 1979. The relationship of visual attention to children's comprehension of television. Child Development, 50, 722-727.

Mandler, G 1975. Mind and emotion. New York, John Wiley and Sons.

McLuhan, M 1964. Understanding media. New York, McGraw-Hill.

Mielke, K and Chen, M 1980. Making contact: Formative research in touch with children. J P Baggaley (ed.) Experimental Research in Televised Instruction, Vol. 5, St. John's University of Newfoundland.

Mueller, C and Donnerstein, E 1981. Film-facilitated arousal and prosocial behaviour. Journal of Experimental Social Psychology, 17, 31-41.

Newcomb, A F and Collins, W A 1979. Children's comprehension of family role portrayals in televised drama: Effects of socio-economic status, ethnicity and age. Development Psychology, 15, 417-423.

Potts, R, Huston, A C and Wright, J C 1982. The effects of television form and violent content on children's attention and social behaviour. Paper presented at the biennial meeting of the South Western Society for Research in Human Development. Galveston, Texas.

Saloman, F 1979. Interaction of Media, Cognition and Learning. San Francisco, Jossey-Bass, Inc.

Schachter, S 1964. The interaction of cognitive and physiological determinants of emotional state. L Berkowitz (ed.) Advances in Experimental Social Psychology, Vol. 1. New York, Academic Press.

Singer, J L and Singer, D G 1980. Television, Imagination and Aggression: A study of pre-schoolers play. Hillsdale, New Jersey, Erlbaum.

Stein, A H and Friedrich, L K 1975. Impact of television on children and youth. E M Hetherington (ed.) Review of Child Development Research, Vol. 5. Chicago, University of Chicago Press.

Tannebaum, P and Zillmann, D 1975. Emotional arousal in the facilitation of aggression through communication. M L Berkowitz (ed.), Advances in Experimental Social Psychology, Vol. 8. New York, Academic Press.

Zillmann, D 1978. Attribution and misattribution of excitatory reactions. J Harvey, W Jakes and R Kidd (eds.) New Directions in Attribution Research, Vol. 2. New York, John Wiley and Sons.

Watt, J H and Krull, R 1974. An information theory measure of television programming. Communication Research, 1, 44-68.

Welch, R L, Huston-Stein, A, Wright, J C and Plehal, R 1979. Subtle sex-role cues in children's commercials. Journal of Communications, 29, 202-209.

Wright, J C and Vlietstra, A G 1975. The development of selective attention: From perceptual exploration to logical search. H W Reese (ed.) Advances in Child Development and Behaviour, Vol. 10. New York, Academic Press.

Wright, J C, Calvert, S L, Huston-Stein, A and Watkins, B A 1980. Children's selective attention to television forms: Effects of salient and informative production features as functions of age and viewing experience. Paper presented at the International Communication Association, Acapulco.

Wright, J C, Huston-Stein, A, Potts, R, Thissen, D, Rice, M, Watkins, B A, Calvert, S, Greer, D and Zapata, L 1980. Formal features of children's TV programmes as predictors of viewership by age and sex: A tale of three cities. Paper presented at the meeting of the South Western Society for Research in Human Development, Lawrence, Kansas.

Wright, J C and Huston, A C 1981. Children's understanding of the forms of television. H Gardner and H Kelly (eds.) Children and the World of Television. San Francisco, Jossey-Bass, Inc.

DETERMINING THE OPTIMAL PLACING FOR YOUNG ADULTS OF ACTIVITIES FOR ETV-PROGRAMMES

R M Yule

The Need for Activities for Use with Educational Television

The importance of activity in the learning situation has often been stressed, but the problem is how can the learner be actively involved in viewing a television programme.

"Local support services will need to be developed, for the amount of learning from instructional television depends at least as much on what happens at the receiving end as what goes into the programme contact and structure" (Richardson, 1980:38).

Wiesner (1983:218) reports that "correspondence materials plus special letters resulted in increased achievement, better attitudes, and greater number of assignments completed". The groups with graded assignments and telephone feedback did better than the control group with correspondence material. In other words, not only is back-up, or support material, important but there should also be a system of feedback so that the learner is aware of progress and learning achievement.

Luskin (1983:49) who examined over 60 educational television programmes states:

> *"All telecourses use textbooks and other assigned readings, the same as campus-based courses. In addition, telecourses require students to obtain and read special study guides that provide instructions regarding assignments, study activities, practice tests, and occasionally, supplemented content ...*

> *In addition, in most courses there are mail quizzes, interaction by telephone and in person with the instructor ..."*

Luskin concludes that telecourses are integrated learning systems, and it is this feature that is very important. The television programme cannot stand on its own.

Forms of Activity

ACTIVITY IS AN ESSENTIAL COMPONENT OF TELEVISION LEARNING

"A widely expressed point of view about television as an instructional medium holds that due to the passivity of the viewer, the medium is virtually powerless to produce learning. There is no question that the medium has limitations in this regard. However, since it is patently obvious that television can and does teach, a more constructive point of view is to examine conceptually ways in which this capability comes about, and operationally, ways in which it may be exploited. The position taken here is that the activity-eliciting potential of the medium, no matter how limited, is nevertheless the chief basis for whatever effectiveness it has" (Palmer, 1972:180).

Thus according to Palmer, television teaching needs a component which involves the learner in active participation. Schramm (1972) concurs with this view. He states:

> *"However, it must be clear, by this time, that the research on content variables comes down very strongly on two points which offer useful guidelines to anyone concerned with programming instructional television or film. One of these is simplicity of presentation, the other active participation by the students in the learning experience" (Schramm, 1972:64).*

Activity is thus pinpointed as an essential component in devising educational television programmes.

INTERACTIVE SYSTEMS

The ideal form of activity will quite obviously be one in which each individual learner is able to interact with the dispenser of information, in this case the television programme, and receive immediate feedback to the response. An era in which this may become possible is fast approaching as "the new frontier for the eighties involves interfacing the computer with video tapes and video discs to create interactive television" (Morris, 1983-84:358). Morris goes on to point out that these innovations have yet to produce significant research results. At this stage there are other simpler systems to be researched.

OVERT AND COVERT RESPONSES

The activity used may only be covert, with no call for an overt response; nevertheless, it is still an activity. Schramm (1972) found from various experiments that overt responses are no better than covert responses. "Most of the differing results on overt vs. covert responding seem to be interpretable on the basis of the kind of material being taught and the rate of presentation" (p 60). It is, however, true that in experimental situations it is easier to control overt responses and actually to observe and measure, whereas covert activity is a non-observable activity.

TYPES OF ACTIVITIES

"Student participation may take many forms. It may consist of repeating a response designed to be learned from the instructional material, answering a question, carrying out some activity described in the material, asking a question about the material, or any of a number of things" (Schramm, 1972:58).

In other words, there are innumerable types of activities that can be devised for use in conjunction with television programmes. Waniewicz (1981), in describing how educational television is used in Ontario, sets out the components of the courses. These include: a learners' guide, reading materials, multiple-choice questions, a computerized system of learning management, newsletters. All of these support materials require some form of activity on the part of the learner.

Some would argue that the material is not as important as the support system itself, and the total teaching-learning system must be well designed and integrated. "Local support services will need to be developed, for the amount of learning from instructional television depends at least as much on what happens at the receiving end as what goes into the programme content and structure" (Richardson, 1980:38).

THE TIMING OF THE ACTIVITY

Activity may be carried out before, during or after viewing an educational television programme, or even possibly at one or more of these stages. It has been found that an appreciable increase in learning resulted when "viewers were given some background written material prior to viewing the segments" (Brown et al, 1974:56). Lundgren (1972:15) maintains, however, that not much activity can be expected during viewing. The literature shows that more emphasis is placed on activity before and after viewing. Durbridge (1981:84) suggests that students "try to watch educational programmes with at least one other person, in order to discuss their ideas about them afterwards, and also be ready to write down their immediate reactions to and thoughts about their content after viewing". Research also indicates that "... after questions are more effective than before questions ... The most effective procedure seems to be to require a response, overt or covert, as soon as possible after the student has been given the information necessary to make the correct response". (Schramm, 1972:61).

Adult Learning and Television

"Experiments in adult educational television have skirted the main issue of the principle of adult learning - how to utilize television to incorporate the modes of learning adjusted best for adult learners" (King, 1979:4).

King goes on to say that what is needed is an integrated learning experience with an emphasis on group discussion. He also sets out five principles for successful adult learning:

* Active is better than passive learning

* Small lesson units, often dealing with a single concept, are desirable

* Correct responses must be recognised and rewarded

* Adult learners pace themselves

* Adults are pragmatic learners and need an individual approach (King, 1979:5)

These principles may, or may not, be accepted. For example, the fourth one that refers to pacing is questionable as research on distance teaching has revealed that pacing of students constitutes a problem. However, what is clear from the work of King and others is that adults need to be actively involved in the learning situation.

Forman and Brown (1978:7) state: "In adult education programs that employ television, facts are generally learned best through print materials. The characteristics of the television medium do not lend themselves to the type of repetition and concentrated study which is often needed to memorize factual materials". Thus, if an experiment is to concentrate on cognitive aspects of learning, it is logical to expect an emphasis on the activity aspects of the television programme in relation to adult learning.

The Experiment

The literature review has established quite conclusively that supplementary activity is necessary for learning to take place with ETV. The problem is at what stage should the activity be introduced. An experiment was done to determine whether there were learning differences associated with the placing of the activity in the programme.

THE GROUPS

The sample was drawn from second year Pretoria College of Education students. The research was done in their free time. Their ages and academic backgrounds were reasonably similar, but some were university students doing a Bachelor of Primary Education degree and others were following a non-academic diploma course. The students were randomly assigned to four groups. Differences in the results could thus be attributed to different treatments of the groups.

SELECTION OF THE EDUCATIONAL TELEVISION TAPE

An Encyclopaedia Britannica television programme on the San Andreas Fault in North America was selected for the experiment. The tape runs for nineteen minutes and contains mainly geographical and geological information. Some of the students may have had background knowledge of the subject content in their school or college studies. This knowledge was determined by the pre-test but discounted by the randomization of the sample selected (see 4.7.1).

THE PRE-TEST AND POST-TEST

The pre-test consisted of 24 items. The same test was used for the pre- and post-tests to permit comparison between the scores.

SUPPLEMENTARY MATERIAL

Supplementary learning material was drawn up in the form of notes on the programme. The students in all four groups were asked to use these notes for the ETV activity.

THE TREATMENT

Table 1: Groups by the Place of the Activity with respect to the ETV-Programme

GROUPS	G	K	L	M
Pre-test	X	X	X	X
Activity before	X			
Activity during		X		
Activity after			X	
Activity before, during and after				X
Post-test	X	X	X	X
Attitude survey	X	X	X	X

As will be seen from Table 1, the pre-test and the post-test, as well as an attitude survey, were administered to all four groups. Group G was given the written notes on the ETV-programme to read through before viewing. Group K had the supplementary notes available during viewing. Group L first viewed the ETV-programme and was then given the notes to read through, whereas Group M had the notes before, during and after viewing to use at any time.

THE OPINION QUESTIONNAIRE

The opinion questionnaire was drawn up for each of the groups. The questions for the groups differed slightly but there were some common elements in the questionnaire such as "Did you learn more from the notes than the programme?".

THE RESULTS - COGNITIVE ASPECTS

- Subjects were randomly assigned to different teaching method groups. This randomization of the subjects discounted the differences in pre-test scores. The experimental treatments were introduced after the pre-tests.

- An Analysis of Variance (ANOVA) was applied to the data to determine whether the groups were samples of the same population, i.e. that the teaching methods did not result in significantly different mean difference scores. The summary of the ANOVA results can be seen in Table 2. The implication of these findings is that there is a significant difference in the teaching methods for the groups overall, at the five percent level.

Table.2: Groups by the Place of the Activity with respect to the ETV-Programme

Source of Variation	D.F.	Sum of Squares	Mean Squares	F-Ratio	F-Probability
Between Groups	3	52,8132	17,6044	2,712	0,0499
Within Groups	86	558,3090	6,4920		
TOTAL	89	611,1222			

An Analysis of Covariance (ANCOVA) was then applied to the post-test scores for the different groups using the pre-test scores as covariate. It can be seen from Table 3, which presents the ANCOVA results, that the F-value for the covariate is significant at the two percent level. The ANCOVA was used to control for the influence of the pre-test scores on the post-test scores. An adjusted mean score was then obtained for each group, and used to determine the level of significance among the groups.

Each pre-test score was adjusted by multiplying it by 0,341 to account for the influence of these scores on the post-test scores.

Table 3: ANCOVA of Post-test Scores by Group with Pre-test Scores as Covariate

Source of Variation	D.F.	Sum of Squares	Mean Squares	F-Ratio	F-Probability
Covariate (Pre-test)	1	31,107	31,107	5,872	0,018
Main Effects (Group)	3	45,037	15,012	2,834	0,043
Residual/Error	85	450,256	19,036	5,297	
TOTAL	89	526,400	5,915		

(Covariate Raw Regression Coefficient (Pre-test) + 0,341)

- Table 4 presents the ANOVA summary showing the effects of different teaching methods on the adjusted difference scores for the four groups of students. When using the sub-program ONE-WAY of the SPSS Package, it is only possible to conduct contrasts and post-hoc tests on the group means. These tests were carried out to investigate specific mean comparisons and post-hoc tests.

Table 4: ANOVA of Adjusted Means of the Difference Scores by Groups

Source of Variation	D.F.	Sum of Squares	Mean Squares	F-Ratio	F-Probability
Between Groups (Adjusted Means)	3	44,9140	14,9713	2,859	0,0416
Within Groups	86	450,3788	5,2370		
TOTAL	89	495,2928			

The F-value were slightly more significant (p0,04) than the F-value for the unadjusted means in Table 1 (p0,049).

Using one-at-a-time tests, it was found that the scores of Group L vs Group M and the scores of Group G vs Group M were significantly different at the 5 percent level. Group K vs Group L was significantly different at the 10 percent level. The post-hoc test results are summarized in Table 5 which shows the results using a Multiple Range Test (LSD) Procedure. Group M is significantly different from Groups G and L.

Table 5: Multiple Range Test - LSD Procedure with ranges for the 0,05 level

GROUP M	GROUP K	GROUP G	GROUP L	Adjusted Means	Groups
				11,7933	M
				12,2115	K
				13,5236	G+
				13,5263	L+

(+) Denotes pairs of groups significantly different at the five percent level. Using Duncan's Procedure, it was found that: Groups K, L, M and Groups M and K did not differ significantly from one another.

The implications of these findings are that the teaching method used for Group M is significantly different from the methods used for Groups L and G at the five percent level, whereas Groups K and L differ at the 10 percent level.

Table 6 presents the means of the group difference scores and pre-test scores.

Table 6: The Means of the Groups Pre-test Score and Difference Scores

GROUP	MEANS OF PRE-TEST SCORES (UNADJUSTED SCORES)	DIFFERENCE SCORE MEANS
L	2,1	12,1429
G	2,2	11,7778
K	2,2	10,7895
M	2,4	10,2174

From Table 6 it can be seen that Group L has the highest mean Difference Score, whereas Group M has the lowest.

* The conclusion that can be drawn from the statistical report is that Group L faired better than the other groups. If the groups are placed in order according to achievement, the order would be L, G, K, M. As stated above, the results were found to be significant at the 10 percent level (in some comparisons even at the five percent level). Using supplementary material after viewing ETV-programme yielded the best results for the ETV-programme viewed. The second most satisfactory method is to let students read supplementary material before viewing. The least satisfactory method of all was to let the students have the material before, during and after viewing.

THE RESULTS - OPINION QUESTIONNAIRE

The responses of the opinion questionnaire are tabulated below.

Table 7: **The opinions of respondents on the usefulness and placing of supplementary materials in respect of the ETV-Programme**

QUESTIONS	RESPONSES OF GROUPS IN PERCENTAGES									
	$G_N=$ 27		$K_N=$ 24		$L_N=$ 21		$M_N=$ 23		TOTAL N=95	
	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO
Did you find the notes useful?	96	4	83	17	95	5	100	0	94	6
Were the notes misleading?	22	78	17	83	14	86	17	83	18	82
Did you learn more from the notes than from the programme?	52	48	33	67	14	86	52	48	39	61
Were you motivated by the notes?	89	11					96	4	-	-
Did the notes help to focus your attention?	78	22					96	4	-	-
Would you have liked to have had the notes to use after viewing?	56	44	75	25					-	-
Were the notes a distraction?			54	46					-	-

Table 7: The opinions of respondents on the usefulness and placing of supplementary materials in respect of the ETV-Programme (continue)

Would you have like to have had the notes to use before viewing?			67	33	57	43			-	-
Was it difficult to follow the notes and the programme at the same time?			83	17					-	-
Were the notes an aid to learning?					95	5			-	-
Were the notes a nuisance during the viewing session?							43	57	-	-
Did you refer to the notes during viewing?							74	26	-	-
Were the notes of greater help before viewing than after viewing?							83	17	-	-
Would you consider using the notes after viewing as of greater use than before viewing?							35	65	-	-

The supplementary material was made available to the groups:

*	Group G -	before viewing
*	Group K -	during viewing
*	Group L -	after viewing
*	Group M -	throughout the process

Comment: Most of the respondents found the notes useful and motivating. However, a meaningful number also found them misleading, distracting and a nuisance during viewing. The respondents as a whole were unsure whether they had learnt more from the notes than from the programme. There was no overall consensus among groups on the ideal placing of the notes in respect of the programme. The difference in opinion could be attributed to differences in learning styles.

Conclusion

More effective learning takes place if supplementary material is used after viewing an ETV-programme, but if material is available before viewing, the viewer may be motivated to learn more. When the viewer is required to take part in an activity during viewing, less learning takes place and the viewer is distracted.

REFERENCES

Aversa, F M and Forman, D C 1978. Issues in the evaluation of educational television programmes. NSPI Journal, Vol. 17, No. 2, March 1978, 16-18.

Baggerley, J, Ferguson, M and Brooks, P 1980. Psychology of the TV Image. Gower Publishing, Farnborough.

Baker, R 1973. Evaluation: the need for standards. Educational Broadcasting International, Vol. 6, No. 1, March 1973, 3-4.

Barbatis, G S 1978. The nature of inquiry and analysis of theoretical progress in instructional television from 1950-1970. Review of Educational Research, Vol. 48, No. 3 1978, 399-414.

Brown, R D, Cavert, C E, Craig, J and Snodgrass, S J 1974. Evaluation of a variety of TV lesson formats for potential adult learners in an open learning system. Educational Technology, Vol. 14, No. 8, August 1974, 55-59.

Carl, D R 1977. Instructional Development in Instructional Television (status and directions). New Jersey, Educational Technology Publications, Englewood Cliffs.

Durbridge, N 1981. The adult learner and educational television. Educational Broadcasting International, Vol. 14, No. 2, June 1981, 82-84.

Everley, J C 1971. The evaluation of continuing education via TV. J A Niemi (ed.) Mass Media and Adult Education. New Jersey, Educational Technology Publications, Englewood Cliffs.

Forman, D C and Brown, L A 1978. The roles of television in adult educational programmes. NSPI Journal, Vol. 17, No. 2, March 19768, 7-8 and 22-25.

Gueulette, D 1980. Television: the hidden curriculum. Lifelong Learning: The Adult Years, Vol. 3, No. 5, January 1980, 4-7.

King, L R 1979. The effective use of television in adult education. Lifelong Learning: The Adult Years, Vol. 2, No. 7, March 1979, 4-7.

Lewy, A 1977. The nature of curriculum evaluation. A Lewy (ed.) Handbook of Curriculum Evaluation. Paris, Unesco.

Lundgren, R 1972. What is a good instructional programme. W Schramm (ed.) Quality in Instructional Television. Honolulu, East-West Center, University Press of Hawaii.

Luskin, B J 1983. Telecourses: 20 myths, 21 realities. Community and Junior College Journal, Vol. 53, No. 8, May 1983, 48-50, 60.

Morris, J D 1983. The Florida study: improvement in student achievement and attitudes through variations in instructional television production values. Journal of Educational Technology Systems, Vol. 12, No. 4, 1983-1984, 357-369.

Palmer, E L 1972. Formative research in educational television production: the experience of the children's television workshop. W Schramm (ed.) Quality in Instructional Television. Honolulu, East-West Center, University Press of Hawaii.

Richardson, P L 1980. Telecommunications and adult learning: What nine projects reveal. M N Chamberlain (ed.) New Directions for Continuing Education. Providing Continuing Education in Media and Technology. San Francisco, Jossey-Bass.

Richardson, P L 1983. Issues in television-centred instruction for adults. Journal of Instructional Development, Vol. 6, No. 3, 1983, 6-13.

Salomon, C and Cohen, A 1977. Television formats, mastery of mental skills and the acquisition of knowledge. Journal of Educational Psychology, Vol. 69, 1977, 612-619.

Schramm, W 1972. What the research says. W Schramm (ed.) Quality in Instructional Television. Honolulu, East-West Center, University Press of Hawaii.

Waniewicz, I 1981. The TV Ontario Academy: the use of television broadcasting and computer-managed learning for adults. Education Broadcasting International, Vol. 14, No. 2, June 1981, 78-81.

Wells, S 1976. Evaluation criteria and the effectiveness of instructional technology in higher education. Higher Education, Vol. 5, No. 3, August 1976, 253-275.

Wiesner, P 1983. Some observations on telecourse research and practice. Adult Education Quarterly, Vol. 33, No. 4, Summer 1983, 215-221.

THE EVALUATION OF ETV-PROGRAMMES IN SOUTH AFRICA: SOME HSRC INITIATIVES

D P van Vuuren and C Herselman

Educational Television in South Africa

ETV in South Africa was initiated in the SABC a few years ago by the formation of two departments for ETV: the Department of Educational Programmes RTV1 and RTV2/3. This development followed the HSRC's investigation and other research on education. These investigations drew attention to the increasing demands on the treasury for education and recommended alternative methods to formal schooling to meet this demand.

Both departments at the SABC have given high priority to evaluation research from the start - which, hopefully, has contributed to the quality and effectiveness of the educational TV programmes currently being broadcast by the different services.

Evaluation Research in South Africa

The introduction of educational TV has promoted evaluation research in South Africa. Two journals published by the SABC contain contributions by researchers and academics of the University of Pretoria (Pretorius, 1985; Sonnekus and Ras, 1985), the University of Fort Hare (Swart, 1985), the University of Potchefstroom for CHE (De Wet, 1985), the Department of Adult Education of the Kwazulu Education and Culture Department (Khanyile, 1985), and the Institute for Communication Research of the HSRC (Groenewald, 1985; Van Vuuren, 1985a, 1985b, 1985c). The SABC's own researchers have also made a valuable contribution to the evaluation of ETV.

The research by Breytenbach (1983) and Fell-Ferreira (1983) on the use of radio and broadcast television in non-formal education was very optimistic: they foresaw exciting possibilities for educational TV.

In this chapter an overview will be given of the HSRC's research initiatives with regard to the evaluation of ETV-programmes. (The HSRC has been doing summative and formative evaluation research for the past two-and-a-half years). A series of fifteen evaluations of nine ETV-programmes will be discussed. Firstly, attention will also be focussed on the development of a methodology for evaluation research, especially the use of micro-computer technology. Thereafter, principles derived from the results will be discussed with reference to the literature on ETV evaluation.

Methodology

BACKGROUND

During 1981 and 1982, the researchers attended two conferences on Experimental Research in TV Instruction at the Memorial University, St John's, Newfoundland, Canada. This encouraged the Media-Effects Division of the Institute for Communication Research of the HSRC to begin formative research with micro-computers. This research effort, which was mainly done for, and in collaboration with, the SABC necessitated a personal approach to evaluation research. Usually long periods elapse before the results of traditional surveys or experimental methods are made known. This did not suit the employer (mostly the SABC): emphasis was placed on quick results, with a view to practical implementation of the evaluation findings. Instead of making use of experts to evaluate the ETV-programmes, the reactions of different target audiences - in one case young children were used as feedback to the employer. (One of Anderson and Field's (1983) seventeen principles is that it is difficult for an adult to know which aspects of the programme a young child gives his attention to).

From interviews with programme planners and producers, the researchers found that the following general expectations for almost all the ETV-programmes needed to be complied with:

- The EFFECTIVENESS of the programme had to be assessed according to the degree in which the programme met its goal ("Will people learn something/become aware of the subject?", or "Will the target audience's behaviour/attitude change with regard to the content?").
- The POPULARITY of a programme ("How many members of the target audience are going to watch? Will they enjoy what they see? Will the theme be real to them?).
- FORMAL FEATURES which influence the effectiveness and popularity of the programme. ("Is the presenter acceptable to the audience? Is the tempo of the presentation correct? Does the audience like the music?").

RESEARCH PROCEDURES

The first step in the research procedure was to make the producer/programme director/SABC programme manager part of the research team. In this way it was possible to meet most of the needs of the director of the programme, and thus bridge the gap identified by Baggaley (1980) and Linton (1982).

Because speed was of utmost importance in the research, and also because the researchers used micro-electronic apparatus, only small samples could be used. The specific target group of each ETV-programme, was the population from which the test samples were drawn.

As the research was exploratory on formative and the summative evaluation, strict experimental methods were not used. Quantitative as well as qualitative data were collected.

The format of the evaluation sessions was usually the same, and consisted of the following elements:

- An introductory talk by the leader of the project.
- Where applicable a pre-implementation questionnaire to determine the viewers' initial knowledge base.
- Screening of the ETV-programme and simultaneous evaluation with the aid of the micro-computer.
- A general programme evaluating questionnaire.
- Group discussion.
- Feedback and discussion with the client and the programme director.

The combining of quantitative (micro-computer) and qualitative (group discussions) methods provided corroboration of the findings.

RESEARCH INSTRUMENTS

* Questionnaires: Cambre (1981) notes that questionnaires have been used for some time to evaluate educational films. Deveroux (1933) developed a check list for grading the production and selection of a movie. The check list comprised a five-point scale ranging from "excellent" to "objectionable".

  In the HSRC study each ETV-programme evaluation used questions on the content of the programme, technical components, and questions about the likes and dislikes of the respondents. These questions dealt with the presenter, the use of music and other variables in the production.

- Micro-computer system: A system known as the Program Evaluation Analysis Computer (PEAC) is mentioned frequently in the literature (see Baggaley, 1985 and Nickerson, 1981). This computer programme is based on Lazerfeld and Stanton's "programme analyser" which was developed in 1940 for the evaluation of radio programmes. It enables the researcher to monitor audience reaction from moment to moment. Initially PEAC was developed for the formative evaluation research done by the Children's Television Workshop (CTW) for the production of "Sesame Street".

  Data collected with this apparatus have high internal validity.

- Group discussion: The qualitative side of the research comprised a group discussion with the target audience. After completing the questionnaires, a structured group discussion was held with the audience. In this discussion all the members of the audience had a chance to give their opinions of the programme.

Results

The results are only tentative. In this chapter we concentrate on data that were collected with the aid of the micro-computer. The data on nine ETV-programmes collected during fifteen sessions from 246 respondents is reported. A summary of the most important findings is given in the tables.

PROGRAMME (EPISODE)	SES-SIONS	TARGET AUDIENCES	N	SCALE	RESULTS FORM MICRO-COMPUTER DATA
"Let's learn Zulu" (1)	3	English speaking Standard 4 children, 10 boys and 10 girls; high socio-economic group, Pretoria.	20 20 20	A = Too slow B = Just right C = Too fast	** Segments where information was presented at too fast a pace, were identified. ** Possible indications that "talking head" and repeating of information elicited "Too slow" responses.
"Time management" (Segments)	1	Randomly selected adults from Johannesburg, English speaking, 3 men, 5 women.	8	A = Clear B = Uncertain C = Not clear	Parts of the segments shown to the audience were not clear to them and those parts were identified and discussed. The audience suggested alternative scenes and contents.
"The world of tomorrow" (1)	1	Standard 9 "Indian" children from Durban. 10 boys and 10 girls.	19	One half: A = Not important B = Little importance C = Fairly important D = Very important Other half: A = Poor B = Not very good C = Fairly good D = Good	** Themes and scenes which were important/not important to the audience were identified: works, competition and modern technology was important to them: "the old days", scenes from rugby, a long scene at a bakery, philosophical discussions and scenes with wild animals were not important. ** If the audience liked a scene, it did not necessarily mean that it was important to them, and vice versa. The graphs "Importance" and "Impression" did not always indicate the same tendencies.

PROGRAMME (EPISODE)	SESSIONS	TARGET AUDIENCES	N	SCALE	RESULTS FORM MICRO-COMPUTER DATA
"Meetings: You take the chair" (1)	1	Adult "Coloureds" from Cape Town and vicinity, 9 men and 12 women.	21	A = Like a lot B = Like a little C = Don't like much D = Don't like at all	Quite a number of preferences/dislikes and tastes were deduced from the graphs: Aversion to violent and disorderly scenes, humour and style of presentation often not applicable. Visual material of Roman buildings and statues was popular. The audience strongly preferred direct, applicable information on the theme.
"The edge" (9)	3	English and Afrikaans-speaking artisans from all population groups, English-speaking white standard 9 children, and English-speaking white adults from Krugersdorp.	20 20 20	A = Like B = Neutral C = Don't like	** Artisans: Difficult problem situations elicited "Don't like" responses. Where problems were solved and more direct information was transmitted, more "Like" responses were found. ** School children: They responded "Neutral" and "Don't like" when the dialogue continued too long, when there was too much technical detail and when one of the actors (the artisan) had to face a difficult situation. Action and the solving of problems drew more positive reaction. The technician who always knew what to do, drew a "Like" response. Boys exhibited extreme reactions, while girls were more moderate. ** Adults: When actors talked too long and covered new information, the adults responded negatively. When scenes changed from talking to action, "Like" responses scored higher. Women reacted with "Dislike" more often than men.

PROGRAMME (EPISODE)	SESSIONS	TARGET AUDIENCES	N	SCALE	RESULTS FORM MICRO-COMPUTER DATA
"Focus on labour relations" (3)	1	Low, middle- and high-level managers from Postal Affairs, Afrikaans-speaking, 16 men and 1 woman.	17	A = Like B = Neutral C = Don't like	** Positive and negative reactions correspond to the goal the producer wanted to reach - the audience reacted negatively when the actors did the "wrong" thing, and positively when the "right" action was taken. ** The presenter was popular and the authority unpopular, according to the graphs. This was verified during the group discussions. ** A number of negative responses towards the end of the programme were recorded possibly due to the repetition of some of the contents.
"Abacus" (1)(4)(9)	3	Afrikaans-speaking standard 5 children, middle socio-economic group, 42 boys and 19 girls from Pretoria.	19 21 21	A = Interesting B = Don't know C = Not interested	** Positive: Colourful scenes/inserts on interesting features and novelties, situations with which the audience could identify and the solving of problems, elicited "Interesting" responses. Humour and comical characters were very popular. ** Negative: Sections where narrative lasted too long without change, situations where they were reminded of school, where questions/problems were put to the audience and situations with which their particular age group couldn't identify (e.g. complicated drawing by an engineer) elicited responses such as "Don't know" and "Not interested".

PROGRAMME (EPISODE)	SESSIONS	TARGET AUDIENCES	N	SCALE	RESULTS FORM MICRO-COMPUTER DATA
"Burns" and "Tuberculosis" (Two health guidance videos in Zulu)	2	Two groups of young black people from Soweto who spoke different languages and all of whom had passed at least standard 6. Aged between 15 and 29 years.	19 19	A = Dislike B = Like	** Positive: "Like" responses increased where (1) family life, children and pretty rooms in a house were pictured, (2) direct information and advice were given, and (3) help, medical treatment and hope was portrayed or discussed. ("Good news"). ** Negative: "Dislike" responses increased where (1) there was talk about possible loss of jobs because of illness, separation from the family, danger and death ("Bad news") and where unacceptable social behaviour was pictured (two drunks).

Discussion

The scale used to measure the responses of the target audience with the micro-computer, was different for the different productions. The following scales were used:

- Too slow
- Just right
- Too fast

- Clear
- Uncertain
- Not clear

- Like
- Neutral
- Don't like

- Interesting
- Do not know
- Not interesting

- Like very much
- Like a little
- Do not like much
- Do not like at all

One half of the group:	Other half of the group:
* Not important	* Poor
* Little importance	* Not very good
* Fairly important	* Fairly good
* Very important	* Good

The following eleven factors were gleaned from the above-mentioned data on which further research should be done. These deductions might be of value to producers of ETV-programmes.

- Tempo: Even for intelligent primary school children, the tempo was sometimes too fast. It is difficult for a producer to judge the tempo correctly.

- School children as well as adults reacted negatively to a "talking head", especially where it was shown for "too long". (Our viewers are becoming more and more "TV-literate").

- A distinction must be made between what the audience enjoys, and what it regards as important.

- Disorderly or violent scenes are not acceptable to all target audiences.

- Lower socio-economic status audiences showed a preference for direct, applicable subject matter and direct didactical methods.

- Difficult problem situations in a training programme elicited negative reactions from apprentices and school children, whereas the solving of problem situations in the programme elicited positive ("Like") responses.

- Continuous talking for long periods elicited negative reactions from school children and adults.

- Males had different reactions to ETV-programmes than did females.

- A popular presenter (on the TV1 channel) elicited positive reactions from the target audiences.

- Humour and the use of comical characters can be used successfully with child audiences.

When studio-shots or close-up shots of a presenter were alternated with interesting and colourful additional material, the reaction was positive.

Although the above-mentioned eleven factors have been drawn (with reservations) from a relatively limited research effort, it is significant that some of these factors comply with research findings abroad. Anderson and Field (1983), for instance, indicated a relationship between humour and the attention of children to television content. In South Africa too, humour seems to "work" in some ETV-programmes for children. Anderson and Field mentioned the difficulty of knowing, intuitively, how to gain the attention of children, and the South African research showed that the producer sometimes wrongly estimated the tempo of the programme, the use of the "talking head" and of dialogue. The four elements identified by Baggaley et al (1980), according to which an audience evaluates a presenter were also observed among South African audiences. (These elements are integrity, mastery, empathy and poise).

Conclusion

Evaluation research of ETV-programmes is complex. The work done for the past two years by the Institute for Communication Research of the HSRC, does not include all the elements of ETV mentioned by Salomon and Del Campo (1981). The research has focused mainly on the content of ETV-programmes, the symbol system of the TV medium and educational-didactic factors experienced by the target audience. In South Africa, contextual research on specific television viewing situations is also very important. The current research initiative is merely a starting point for further research in this field.

REFERENCES

Anderson, D R and Field, D E 1983. Children's attention to television: implications for production. M Meyer (ed.) Children and the Formal Features of Television. Saur München.

Baggaley, J P, Ferguson, M and Brooks, P 1980. Psychology of the TV Image. Gower Westmead, Farnborough.

Baggaley, J P 1985. Formative evaluation of educational television. Educational Programmes, RTV1, 1(1). Johannesburg, SABC.

Breytenbach, L L 1983. The use of Radio and Broadcast Television in Formal Education. Unpublished M Ed-script. Johannesburg, Rand Afrikaans University.

Cambre, M A 1981. Historical overview of formative evaluation of instructional media products. Educational Communication and Technology - a Journal of Theory, Research and Development, 29(1), 3-25.

Devereux, F L 1933. The Educational Talking Picture. Chicago, University of Chicago Press.

De Wet, J J 1985. How do you write a television programme? The experiences of a novice. Educational Programme RTV 1, 1(1). Johannesburg, SABC.

Fell-Ferreira, A 1983. The use of Radio and Broadcast Television in non-formal Education. Unpublished M Ed-script. Johannesburg, Rand Afrikaans University.

Groenewald, H J, Bekker, S J, Bornman, E and Van Aswegen, A 1985. Social learning needs within certain groups of blacks in South Africa. RTV 1985. Informal Educational Broadcasting in Perspective, 1(1), 35-42. Johannesburg, SABC.

Khanyile, E 1985. Television as an adult education medium in the modernization of Zulu culture. RTV 1985. Informal Educational Broadcasting in Perspective, 1(1), 46-49. Johannesburg, SABC.

Linton, J M 1982. Teaching production research and design: The interface of theory and practice. J P Baggaley and P Janega. Proceedings of the Fourth International Conference on Experimental Research in Televised Instruction, Vol. 4. Concordia University.

Nickerson, R 1981. Second generation electronic testing technology applied to television evaluation. J P Baggaley and P Janega. Proceedings of the Fourth International Conference on Experimental Research in Televised Instruction, Vol. 4. Concordia University.

Pretorius, J W M 1985. Compensatory education: a research need of education radio and television. RTV 1985. Informal Broadcast-Education in Perspective, 1(1), 9-16. Johannesburg, SABC.

Sonnekus, M C H and Ras, B 1985. The possible contribution of pre-school television programmes to learning readiness. RTV 1985. Informal Broadcast-Education in Perspective, 1(1), 22-28. Johannesburg, SABC.

Swart, F A 1985. Evaluation of informal educational radio programmes in Xhosa. RTV 1985. Informal Broadcast-Education in Perspective, 1(1), 50-56. Johannesburg, SABC.

Van Vuuren, D P 1985a. The challenge of the electronic mass media. Educational Programme RTV 1, 1(1), 9-19. Johannesburg, SABC.

Van Vuuren, D P 1985b. Formative research evaluation of educational television programmes for black pre-schoolers. RTV 1985. Informal Broadcast-Education in Perspective, 1(1), 29-34. Johannesburg, SABC.

Van Vuuren, D P and Herselman, C 1985c. Formative evaluation of an educational programme "The Edge". Educational Programmes RTV 1, 1(1), 64-74. Johannesburg, SABC.

CONCLUDING COMMENTS

A Lewy

The articles contained in this book represent a scholarly attempt to deal comprehensively with problems related to evaluating educational television programmes. Quite likely it is the first book devoted fully to this topic.

The aims of the book have been: the systematization of knowledge available in this field; the documentation of much-needed new knowledge; and the "transmission of a systematised body of knowledge to interested readers".

The target audiences of the book are: educational leaders responsible for policy decisions on the production of ETV-programmes or responsible for commissioning productions; directors and producers of ETV-programmes; professional evaluators; and those people who face the task of advising on the selection of programmes to suit specific needs.

The evaluation of ETV-programmes is conceived as a special aspect of the broader area of programme evaluation. Criteria used for programme evaluation in general have constituted a starting point for specifying criteria for ETV-programmes.

The repertoire of programme evaluation criteria has been enriched in two different ways. Firstly, attempts have been made to identify unique features of ETV-programmes and to use them as a basis for deriving additional criteria. In defining the exact boundaries of ETV-programme evaluation, the authors have subtly differentiated between criteria appropriate for evaluating entertainment programmes, and criteria appropriate for evaluating ETV-programmes.

In doing this, they have dispelled the misconception that one can define the domain of ETV-programme evaluation criteria by putting together all the criteria used for evaluating educational programmes and TV programmes. Indeed, the domain occupies only the area where the above-mentioned two criterion areas meet.

Secondly, the repertoire of criteria has been enriched by identifying the common features of programme evaluation on the one hand, and evaluation activities in a broad range of life situations, on the other hand. The book has examined features of literary criticism, art criticism, investigative journalism, etc. and shown to those employed in the field of programme evaluation the similarity between evaluation criteria used in these fields.

The systematic treatment of evaluation criteria in this book constitutes a major contribution to evaluation theory in general: it illustrates common features of evaluation activities of any kind, and across many aspects of life and also calls attention to the uniqueness of evaluation activities in highly specific domains of human endeavour. In the last three decades, we have witnessed a high level of differentiation in the field of evaluation research. Programme Evaluation has emerged as a new field of specialization, and has become strongly detached from the main stream of research in the discipline known as Measurement and Evaluation. Now we are witnessing a trend towards further differentiation in the domain of programme evaluation. The terms Curriculum Evaluation, Evaluation of Social Actions, Implementation Evaluation, Evaluation of ETV-programme denote emerging areas of professional activity, with the domain of Programme Evaluation, and their further development may lead toward additional fragmentation in this field. In the face of such trends one cannot overestimate the importance of weaving interpretive threads and pointing out common features of evaluation research and study in many areas of life. This book has tried to do this.

Another important feature of the book is also the authors' attempts to address the management aspects of evaluation. They have suggested that evaluation should be done as a co-operative endeavour between all stakeholders involved in a particular ETV-programme. The collaboration between persons who fulfil different roles in the process of programme production is likely to ensure that evaluation activities will focus on important aspects of the programme, and that evaluation results will be available in due time. In this way one may increase the chances of the evaluation results being used in the process of decision making.

Empirical studies carried out in South Africa contained in this book have touched upon the needs and characteristics of the local audiences of ETV-programmes. In these studies preference has been given to pre-school children and adult viewers.

No empirical studies on the use of ETV in schools are contained in this book: at the time of preparation, the educational programmes of the South African Broadcasting Service were serving uniquely the needs of informal education. No ETV-programmes have been developed or transmitted by the SABC for the formal educational system. Consequently, problems related to using ETV-programmes as a support instrument, or as a supplementary device in the school have not yet emerged in South Africa. Nor have problems of integrating ETV-programmes into the routine schedule of the school emerged. Systematic evaluation of ETV-programmes used in schools within the framework of their routine operation will be possible only at a later stage when sufficient experience has been accumulated in South Africa. This theme may be the focus of another book.

Printed and Bound by Perskor Book Printers – (B9923)